

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

BEST AVAILABLE COPY

STAFF USE ONLY

Date completed: 12-09-03
Searcher: Beverly C4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site
____ STIC
____ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ ☒ Other CGN

THIS PAGE BLANK (USPTO)



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 109333

To: **Manjunath N Rao**
Location: **cm1/10a11 & 10d01**
Art Unit: **1652**
Wednesday, December 10, 2003

Case Serial Number: 09/980729

From: **Beverly Shears**
Location: **Biotech-Chem Library**
CM1-1E05
Phone: **308-4994**

beverly.shears@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

109333

From: Rao, Manjunath N.
Sent: Monday, December 01, 2003 9:36 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/980729

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Date: 12-1-03

Please search the following as soon as possible for application with serial number

09/980,729

1. SEQ ID NO: 10 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 5 against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Mail Box in 10D01
Crystal Mall 1, USPTO.

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:55 ; Search time 18 Seconds
(without alignments)
1382.063 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSMKMTSALLILQLSCYFSS.....KCFLLSCQKFKTRKIEKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_A1.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2602.5	93.0	528	1	UDBA_HUMAN
2	2369.5	84.7	528	1	UDBJ_MACFA
3	2330	83.3	529	1	UDB7_HUMAN
4	2297	82.1	529	1	UDB9_MACFA
5	2292	81.9	529	1	UDBE_HUMAN
6	2281.5	81.5	528	1	UDBA_HUMAN
7	2253	80.5	529	1	UDBS_HUMAN
8	2169.5	77.5	530	1	UDBE_HUMAN
9	2158.5	77.1	530	1	UDBF_HUMAN
10	2100.5	75.1	530	1	UDBK_MACFA
11	1991.5	71.2	530	1	UDBE_RABIT
12	1973.5	70.5	529	1	UDBI_RAT
13	1938.5	69.3	523	1	UDBG_RABIT
14	1938	69.3	531	1	UDBD_RABIT
15	1930.5	69.0	530	1	UDBC_RAT
16	1904.5	68.1	530	1	UDBS_MOUSE
17	1894.5	67.7	530	1	UDB3_RAT
18	1856.5	66.4	530	1	UDB2_RAT
19	1840.5	65.8	530	1	UDB6_RAT
20	1777.5	63.5	530	1	UDB8_RAT
21	1761.5	63.0	527	1	UDAI_RAT
22	1593	56.9	502	1	UDCI_RABIT
23	1201.5	42.9	531	1	UDIS_RAT
24	1196.5	42.8	533	1	UD12_RAT
25	1188.5	42.5	533	1	UD12_MOUSE
26	1158	41.4	535	1	UD11_MOUSE
27	1137.5	40.7	531	1	UD13_RAT
28	1137.5	40.7	533	1	UD11_HUMAN
29	1135.5	40.6	535	1	UD15_RAT
30	1128	40.3	534	1	UD15_HUMAN
31	1126	40.2	534	1	UD16_HUMAN
32	1124	40.2	534	1	UD16_HUMAN
33	1120.5	40.0	530	1	UD18_HUMAN

34	1120.5	40.0	530	1	UD19_HUMAN
35	1109.5	39.7	532	1	UD14_RABIT
36	1099.5	39.3	530	1	UD1A_HUMAN
37	1098.5	39.3	530	1	UD12_HUMAN
38	1094.5	39.1	530	1	UD17_HUMAN
39	1094	39.1	520	1	UD17_MOUSE
40	1081.5	38.7	532	1	UD16_HUMAN
41	1078.5	38.5	531	1	UD17_RAT
42	1077.5	38.5	530	1	UD18_RAT
43	1066	38.1	531	1	UD16_MOUSE
44	1045.5	37.4	529	1	UD16_RAT
45	1045	37.3	531	1	UD16_RABIT

ALIGNMENTS

RESULT 1

UDBA_HUMAN STANDARD; PRT; 528 AA.
AC P06133; 060731; 060867; 075614; P36538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UDP-glucuronosyltransferase 2B4 precursor, microsomal (EC 2.4.1.17)
DE (UDPgt) (Hydroxycholelic acid) (HUG25) (UDPgtH-1).
GN UGT2B4 OR UGT2B11.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87241362; PubMed=3109396;
RA Jackson M.R., McCarthy L.R., Harding D., Wilson S., Coughtrie M.W.H.,
RT Burchell B.;
RT "Cloning of a human liver microsomal UDP-glucuronosyltransferase
RT cDNA.";
RL Biochem. J. 242:581-588(1987).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93326164; PubMed=8333863;
RA Jin C.-Y., Miners J.O., Lillywhite K.J., McKenzie P.I.;
RT "cDNA cloning and expression of two new members of the human liver
RT UDP-glucuronosyltransferase 2B subfamily.";
RL Biochem. Biophys. Res. Commun. 194:496-503(1993).
[3]
RP SEQUENCE FROM N.A., VARIANT GLU-458, AND CHARACTERIZATION.
RX MEDLINE=99303261; PubMed=10376768;
RA Levesque E., Beaulieu M., Hum D.W., Belanger A.;
RT "Characterization and substrate specificity of UGT2B4 (B458): a
RT UDP-glucuronosyltransferase encoded by a polymorphic gene.";
RL Pharmacogenetics 9:207-216(1999).
[4]
RP SEQUENCE FROM N.A., AND VARIANTS LEU-109 AND LEU-396.
RA McKenzie P.I.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA Riedy M., Miller A.;
RT "Genomic organization and structure of the UGT2B gene complex at human
RT chromosome 4q31.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: UDPGTs are of major importance in the conjugation and
subsequent elimination of potentially toxic xenobiotics and
endogenous compounds. This isozyme is active on polihydroxylated
estrogens (such as estradiol, 4-hydroxyestrone and 2-hydroxyestradiol)
and xenobiotics (such as 4-methylumbelliferone, 1-naphthol, 4-
nitrophenol, 2-aminophenol, 4-hydroxybiphenyl and menthol). It is
capable of 6 alpha-hydroxyglucuronidation of hydroxycholelic acid.
-!- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

THIS PAGE BLANK (USPTO)

[illegible]

Matches 492; conservative 14; Mismatches 22; Indels 1; Gaps 1

100

61 SISFDPNPSTLKFEVYPVSLTKTEFEDI IKOLVKRWAEI.PKDTMFSYFSOVQETMMTEN 120

120 I M W I E N I M I E I N Q U A R T E L I F S I S V Q E I M W I E N

121 DILRFECKDIVSNKLMKLQESRFDVVLADAVFPFGELLAELKIPFVYSLRFSFGYAI 180

Db 181 EKHSGLFPSPYVPVMSELSDDMTFIERVKNMIVLYEEFWFOIEDMKKWDQFYSEVL 240

241 CBPBTI CEMAKADITW I DUVNDEOBNYU ? BUREMUGA HOUTA WTA TUNUTU

301 ENGVVFLGSMVSNSEERANVIASALAKIPQKVLWRFDGKNKPDTLGLNTRLKWPON 360

1

RESULT	ID	UDBU_MACFA	STANDARD;	PRT;	528 AA.
2	AC	Q92T55;			
DT	30-MAY-2000	(Rel. 39,	Created)		
DT	30-MAY-2000	(Rel. 39,	Last sequence update)		
DT	16-OCT-2001	(Rel. 40,	Last annotation update)		
DE	UDP-glucuronosyltransferase 2B19 precursor,				
NB	UGT2B19.				

OC Macaca fascicularis (leaf eating macaque) (Cynomolgus monkey)
OC Macaca discolor (leaf eating macaque) (Cynomolgus monkey)
OC Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoinae; Macaca
OX NCBI_TaxId=9541;

RP SEQUENCE FROM N.A.

RX MEDLINE=99203465: PubMed=10102998.
RC 1 ISSUE=Liver, and prostate;
RC 1

RA Belanger G., Barbier O., Hum D.W., Belanger A.;
PT "Molecular cloning and characterization of a novel

RT steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates

Eur. J. Biochem. 260:701-708(1999).

CC -!- FUNCTION: UDEGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC

ENDOGENOUS COMPOUNDS. THIS ISOZYME DISPLAYS ACTIVITY TOWARD SEVERAL CLASSES OF XENOGENIC SUBSTRATES.

METHYLLUMBERONE, P-NITROPHENOL, 1-NAPHTHOL, P, P' -BIPHENOL,

HYDROXY POSITIONS OF STEROIDS.

-1- FUNCTION: TO THE FORMATION OF ANDROGEN GLUCURONIDE IN
CC EXTRAHEPATIC TISSUES SUCH AS THE PROSTATE
CC

CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

CC -1- SUBCELLULAR LOCATION: Microsomal (By similarity).

CC KIDNEY, PANCREAS, BRAIN, CEREBELLUM, MAMMARY GLAND AND EPIDIDYMIS

NOT EXPRESSED IN SMALL INTESTINE, SPLEEN, BLADDER, ADRENAL GLAND AND TESTIS.

CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collabor

the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed, reuse by and for commo

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>)

DR InterPro; IPR002213; UDP gluc trans.

```
DR PROSITE: PS00375: HDPGT: 1
DR PROSITE: PS00201: UDPGT: 1
```

Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal

FT	SIGNAL	1	21	POTENTIAL.

FT	493	513	POTENTIAL.
TRANSMEM			002 - SUBCUNONOSILININSTR ENAGE 2B17.

SEQUENCE	AA:	MMW:	3BED2AE714A27AEF	CRC64:
F1 CARBOHYD	315	315	N-LINKED (GLCNAC: . . .)	(POTENTIAL).
S0 SEQUENCE	528	AA:	60741	MMW:

THIS PAGE BLANK (USPTO)

QY	901	GAAGGAAATGAAAGAGTTGTCAGAGCTCTGGAGAAAAATGGTGTTGGTGTCTCT	960
Db	901	GAAGGAAATGAAAGAGTTGTCAGAGCTCTGGAGAAAAATGGTGTTGGTGTCTCT <td>960</td>	960
QY	961	GGGGTCGATGTCAGTAAACGTCAGAGAAAGGSCCAATGTAATTCATCAGCCCTTGC	1020
Db	961	GGGGTCGATGTCAGTAAACGTCAGAGAAAGGSCCAATGTAATTCATCAGCCCTTGC	1020
QY	1021	CAAGATCCCAAAAAGGTTCTGGAGATTGATGGAAATAAACAGATCTTTAGACT	1080
Db	1021	CAAGATCCCAAAAAGGTTCTGGAGATTGATGGAAATAAACAGATCTTTAGACT	1080
QY	1081	CAATACCTCGGCTGTACAGTGAATACCCCAAGATGATCTTTCTTGTCATCCCAAAACCA	1140
Db	1081	CAATACCTCGGCTGTACAGTGAATACCCCAAGATGATCTTTCTTGTCATCCCAAAACCA	1140
QY	1141	AGCTTTATCATCTCATGTGGAAATGAATGGAGTCTATGAAGCTATTTACATGGGCTCC	1200
Db	1141	AGCTTTATCATCTCATGTGGAAATGAATGGAGTCTATGAAGCTATTTACATGGGCTCC	1200
QY	1201	TATGGTGGAGTCCCAATTTGGTATCATGCTGTATCATAGCTCCACATGAAGGCCAA	1260
Db	1201	TATGGTGGAGTCCCAATTTGGTATCATGCTGTATCATAGCTCCACATGAAGGCCAA	1260
QY	1261	AGGAGCAGCTGTAGAAATTAACCTTCAAACTATGACAAAGCAAGATTTACTGAGGCTTT	1320
Db	1261	AGGAGCAGCTGTAGAAATTAACCTTCAAACTATGACAAAGCAAGATTTACTGAGGCTTT	1320
QY	1321	GAGAAACAGTATTCGATCTCTTATTAAGAGATGCTATGAGATTATCAAGAAATTC	1380
Db	1321	GAGAAACAGTATTCGATCTCTTATTAAGAGATGCTATGAGATTATCAAGAAATTC	1380
QY	1381	CCAGATCAACCTGTAAAGCCCTAGATCGAGCAGTTCGAGTCGAGTTTGCATGCG	1440
Db	1381	CCAGATCAACCTGTAAAGCCCTAGATCGAGCAGTTCGAGTCGAGTTTGCATGCG	1440
QY	1441	CCACAAAGAGCCAAAGCAGCTGCATAGCTGCCATGACCTCACTCTGGTTCCAGACTA	1500
Db	1441	CCACAAAGAGCCAAAGCAGCTGCATAGCTGCCATGACCTCACTCTGGTTCCAGACTA	1500
QY	1501	CTCTATGATGTGATTTGGGTTCTCTGCTGACCTGTGTGGCAATGCTATATTTCTGTTCAC	1560
Db	1501	CTCTATGATGTGATTTGGGTTCTCTGCTGACCTGTGTGGCAATGCTATATTTCTGTTCAC	1560
QY	1561	AAATATGTTTTTATTTCTCTGCTCAAAATTTAATAAATCTAGAAAGATAGAAAGAGGGA	1620
Db	1561	AAATATGTTTTTATTTCTCTGCTCAAAATTTAATAAATCTAGAAAGATAGAAAGAGGGA	1620
QY	1621	ATAGATCTTTCCAAATTCAGAAAGACCTG	1650
Db	1621	ATAGATCTTTCCAAATTCAGAAAGACCTG	1650
RESULT 2			
LOCUS	HSNJ5162	2079 bp	linear
DEFINITION	HSNJ5162	2079 bp	linear
ACCESSION	AJ005162		
VERSION	AJ005162.1	GI:3135024	
KEYWORDS	UDP-glucuronosyltransferase; UGT2B4 gene.		
SOURCE	Human sapiens (human)		
ORGANISM	Human sapiens		
REFERENCE	1	Jackson, M.R., McCarthy, L.R., Harding, D., Wilson, S., Coughtrie, M.W. and Burchell, B.	
AUTHORS			
TITLE			
JOURNAL	Biochem. J.	242 (2), 581-588	(1987)
PUBMED	87241362		
PMED	3109396		

THIS PAGE BLANK (USPTO)


```

REFERENCE
AUTHORS      2 Ritter,J.K., Chen,F., Sheen,Y.Y., Lubet,R.A. and Owens,I.S.
TITLE        Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
              log differences in activity toward parallel substrates including
              hydroxycholeic acid and certain estrogen derivatives
JOURNAL      Biochemistry 31 (13), 3409-3414 (1992)
MEDLINE      92207964
PUBMED       1554722
REFERENCE     3 (bases 1 to 2079)
AUTHORS      Ritter,J.K.
TITLE        Direct Submission
JOURNAL      Submitted (08-MAY-1998) Ritter J.K., Pharmacology and Toxicology,
              Virginia Commonwealth University, P.O. Box 980613, Medical College
              of Virginia, Richmond, Virginia, 23112-0613, USA
FEATURES
SOURCE
1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cdname="UDPGLT1"
/clone_lib="Lambda bacteriophage gtl1"
/tissue_lib="Liver"
1..2079
/gene="UGT2B4"
34..1620
/gene="UGT2B4"
/EC_number="2.4.1.17"
/function="transfers glucuronic acid to hydroxycholeic
acid and estrogen catechols"
/codon_start=1
/evidence=experimental
/product="UDP-glucuronosyltransferase"
/protein_id="CA06396.1"
/db_xref="GI:3135025"
/db_xref="SPTREMBL:O60867"
/translation="MSMKWTSALLILQLSCYFSSGCGKLVLPTEFSHMNITKILID
ELNORGHVTVYIASASISPDNSPSTLKEFVYPSVLTKEFEEDIIKOLVRYAELEPKR
DLFWFSYQVOELIMTFNDILIRKFEKDVNSKKLKKLQESRPVYLDVAKNIPFGCELL
AELLKTFPVSLRFPSPGYAIEKHSGLLPSPSYFVNVNELSDQKTFIERKNMITYL
YEFEMFQI FDMKKMDQFYSEVIGRPTLSEFMKADWLIRYWDFOFPHLPNVEFE
VGLKRCRPAKPLPKMEEFVQSSGNGVVFVSLGSMVSNTESEKAVNYASLAKI POK
VLMRFDQGNKPTLIGNTRLYKWI PONDLLGHPKTRPAFTTGGANGIYKAIYHGI PMVQ
VLPLEADQDNITAHMKAKGAAYSLDIFHTMSDTDLNALKTVINPILYKNAKLSRIHH
DQVPELRNAVFWIEFVWRHKGAKLRVAADLTWFOYHSLDVTGFLACATVIFITII
TKLCLFCVMKEFVTKGGRKD"
546
conflict
/gene="UGT2B4"
/citation=[1]
/replace="g"
548..549
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1187..1188
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1190
/gene="UGT2B4"
/citation=[1]
/replace="c"
1192
/gene="UGT2B4"
/citation=[1]
/replace="a"
1977..1978
/gene="UGT2B4"
/citation=[1]
/replace="cga"
2068
/gene="UGT2B4"
/citation=[1]
/replace="g"

```

BASE COUNT	634 a	396 c	436 g	613 t	
ORIGIN					
Query Match	87.5%; Score 1444.4; DB 9; Length 2079;				
Best Local Similarity	93.3%; Pred. No. 0;				
Matches 1521, Conservative	0; Mismatches 106; Indels 3; Gaps 1;				
Qy	2	GCACCTGAAAAACAAGCATTCATTCAGATGATGCTATGTAAGAAATGACCTTGAGCTCTT	61		
Db	1	GCAACTGAAAAACAAGCATTCATTCAGATGATGCTATGTAAGAAATGACCTTGAGCTCTT	60		
Qy	62	CTCCTGATACAGCTGACCTGTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT	121		
Db	61	CTGCTGATACAGCTGACCTGTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT	120		
Qy	122	CCCAAGAAATTCAGCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG	181		
Db	121	CCCAAGAAATTCAGCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG	180		
Qy	182	GCTCATGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	241		
Db	181	GCTCATGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	240		
Qy	242	TCTACTCTTAATTTGAAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT	301		
Db	241	TCTACTCTTAATTTGAAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT	300		
Qy	302	ATCAAGACAGCTGTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG	361		
Db	301	ATCAAGACAGCTGTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG	360		
Qy	362	TCAACAAGTACAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG	421		
Db	361	TCAACAAGTACAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG	420		
Qy	422	ATAGTTTCAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT	481		
Db	421	ATAGTTTCAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT	480		
Qy	482	GCAGATGCTGTTTCCCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	541		
Db	481	GCAGATGCTGTTTCCCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	540		
Qy	542	TACAGCTCCGCTTCT	601		
Db	541	TACAGCTCCGCTTCT	600		
Qy	602	CCTCCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	661		
Db	601	CCTCCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	660		
Qy	662	AGGGTAAAAAATATGATCTATGCTTTATTTTGAATTTTGGTCCAAATATTTGACATG	721		
Db	661	AGGGTAAAAAATATGATCTATGCTTTATTTTGAATTTTGGTCCAAATATTTGACATG	720		
Qy	722	AAGAATGGATCAGTTCACAGTGAAGTTCAGGAAGCCACTACGTTATCTGAGACA	781		
Db	721	AAGAATGGATCAGTTCACAGTGAAGTTCAGGAAGCCACTACGTTATCTGAGACA	780		
Qy	782	ATGGCAAAAAGTACATATGAGTTATTCGAAAATACAGTGGATTTTCAATTTCTGACCCA	841		
Db	781	ATGGCAAAAAGTACATATGAGTTATTCGAAAATACAGTGGATTTTCAATTTCTGACCCA	840		
Qy	842	CTCTTACCAATGTTGATTCGTTGAGAGATCTCACTGAAACCTGCCAAACCCCTACCG	901		
Db	841	CTCTTACCAATGTTGATTCGTTGAGAGATCTCACTGAAACCTGCCAAACCCCTACCG	900		
Qy	902	AAGGAATGGAAGTTCACAGACCTGAGGAATAAGTGTGTGTGTGTGTGTGTGTGTGTGT	961		
Db	901	AAGGAATGGAAGTTCACAGACCTGAGGAATAAGTGTGTGTGTGTGTGTGTGTGTGTGT	960		
Qy	962	GAGTCAGATGTCAGTAAACGTCAGAAAGAGGCAATGTAATTCATCAGCCCTTGC	1021		

THIS PAGE BLANK (USPTO)

Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-simg.stanford.edu>
Contact: (Dickson, Mark) mdc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAL Plate: 37 Row: c Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10863940
Location/Qualifiers
1. 2129
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="Locusid:7363"
/db_xref="taxon:9606"
/clone="MGC:22623 IMAGE:4709436"
/tissue_type="Liver"
/clone_lib="NH MGC_76"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
45. 1631
/codon_start=1
/product="UDP glycosyltransferase 2 family, polypeptide B4"
/protein_id="AAH26264.1"
/db_xref="GI:20070776"
/translation="MSMKWTSALLLIOLSCYSSGCGKLVLPETSRHMNIKITLLILVLRGHEVAVLWTSASISFDPNPSPLTFEYVSLTKEFEDIKOLVIRMAELPPELDTFMSRVOYEIMMTFNDILRKRCXDIVSKKLMKLOESRFPVLADAVPPELHBLIKIPVYSLRSPGAIIEKHGGGLPFPSPYAPVYMSESDMTFERKXNIVYVFEPEFQIFDMKMDQFYSEVLGRPTLSEMAKADWILKNYDFOPPHLLPVEEVLGHCXKAPKPLPEMEEFVQSSENGVSFSLMSVNSSEBANVASLAKIPOVLRMFEDPDGLGLNLRILYKMIQONDLGHPKTRAFTHGAGNIVKAIYHGIIPMWVLEPADPDNPIAHMKAGAAVSLDFHTMSSDILNALKTVINDLYKENAMKLSIHHDDPKRLEDAVFMEIEFPMRHGAHLVAADHLWFOHSLDVGLACAVATVIFITKLCFCYMKRVTRKGRKRD"

BASE COUNT 672 a 398 c 441 g 618 t
ORIGIN

Query Match 87.5%; Score 1443.8; DB 9; Length 2129;
Matches 1521; Conservatvie 0; Mismatches 107; Indels 3; Gaps 1

1 AGCAACTGGAAAAAGACGATTCATTGCGATCAGGATGCTCTATGAATGCACTTCAGCTCT 60
11 AGCAACTGGAAAAAGACGATTCATTGCGATCAGGATGCTCTATGAATGCACTTCAGCTCT 70

61 TCTCCTGATACAGCTGAGCTGTTACTTTAGCTCTGGGAGTTGTGGAAGGTCCTGCTG 120
71 TCTGCTGATACAGCTGAGCTGTTACTTTAGCTCTGGGAGTTGTGGAAGGTCCTGCTG 130

121 GCCCAAGAAATTCAGCCACTGGAGTAATATAAGACATCCGATGAGTAACCTGTCAGAG 180
131 GCCCAAGAAATTCAGCCACTGGAGTAATATAAGACATCCGATGAGTAACCTGTCAGAG 190

181 AGGTCATAGGTCATGTAATGGATCTTCAGCTTCATTTCTTTGATCCCAACAGCCC 240
191 AGGTCATAGGTCATGTAATGGATCTTCAGCTTCATTTCTTTGATCCCAACAGCCC 250

241 ATCTACTCTTAAATTTGAAGTTTATCCTGTATCTTTAACTTAAACTGAGTTTGAGAGAT 300
251 ATCTACTCTTAAATTTGAAGTTTATCCTGTATCTTTAACTTAAACTGAGTTTGAGAGAT 310

301 TATCAAGAGCTGTGTTAAGATAGGCGAAGCTCCAAAGACATTTGGTCATATTT 360
311 TATCAAGAGCTGTGTTAAGATAGGCGAAGCTCCAAAGACATTTGGTCATATTT 370

361 TTCACAAAGTACAAGAAATCATGTGGACATTTAATGACATACTTGAAGAGTTCTGTAAAGA 420
371 TTCACAAAGTACAAGAAATCATGTGGACATTTAATGACATACTTGAAGAGTTCTGTAAAGA 430

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:47:25 ; Search time 21 Seconds
(without alignments)
1065.631 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798
Sequence: 1 MSKMTSALLLIOLSCYFSS.....KCLFSCQKFNKTRKIERRE 529

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2602.5	93.0	528	4	US-09-356-806-8
2	2322	83.0	524	4	US-09-356-806-40
3	2169.5	77.5	530	3	US-09-180-852-2
4	2158.5	77.1	530	4	US-09-356-806-113
5	1874.5	67.0	454	4	US-09-813-918-2
6	1243	44.4	288	4	US-09-813-918-3
7	1196.5	42.8	531	5	PCT-US92-00282-6
8	1130.5	40.4	533	5	PCT-US92-00282-3
9	1126	40.2	534	5	PCT-US92-00282-4
10	1064	38.0	531	5	PCT-US92-00282-5
11	1045.5	37.4	529	5	PCT-US92-00282-7
12	833	29.8	197	4	US-09-813-918-4
13	765.5	27.4	245	4	US-09-305-856B-18
14	403	14.4	515	3	US-08-942-012B-32
15	393	14.0	129	4	US-09-370-838-36
16	379	13.5	488	3	US-08-942-012B-29
17	379	13.5	488	3	US-08-942-012B-30
18	374	13.4	460	3	US-08-942-012B-33
19	372	13.3	288	4	US-09-305-856B-2
20	372	13.3	288	4	PCT-US92-00282-19
21	362.5	13.0	289	4	US-09-305-856B-6
22	360.5	12.9	289	4	US-09-305-856B-8
23	360.5	12.9	289	4	PCT-US92-00282-13
24	358.5	12.8	289	4	US-09-305-856B-4
25	358.5	12.8	289	4	PCT-US92-00282-15
26	354.5	12.7	289	5	PCT-US92-00282-11
27	346	12.4	310	4	US-09-305-856B-14

28	338	12.1	515	3	US-08-942-01	Sequence 24, Appl
29	321.5	11.5	317	4	US-09-305-856B-10	Sequence 12, Appl
30	316	11.3	287	4	US-09-305-856B-10	Sequence 10, Appl
31	306.5	11.0	286	5	PCT-US92-00282-9	Sequence 9, Appl
32	297.5	10.6	98	5	PCT-US92-00282-26	Sequence 26, Appl
33	296.5	10.6	253	4	US-09-305-856B-16	Sequence 16, Appl
34	294.5	10.5	506	3	US-08-942-012B-26	Sequence 26, Appl
35	287	10.3	506	3	US-08-942-012B-25	Sequence 25, Appl
36	287	10.3	506	6	5180581-2	Patent No. 5180581
37	271	9.7	58	2	US-08-466-583-9	Sequence 9, Appl
38	271	9.7	58	2	US-08-265-427-9	Sequence 9, Appl
39	271	9.7	58	5	PCT-US95-07820-9	Sequence 9, Appl
40	261	9.3	493	3	US-08-942-012B-28	Sequence 28, Appl
41	257	9.2	52	2	US-08-466-583-7	Sequence 7, Appl
42	257	9.2	52	4	US-08-265-427-7	Sequence 7, Appl
43	257	9.2	52	4	PCT-US95-07820-7	Sequence 7, Appl
44	250.5	9.0	489	3	US-08-942-012B-31	Sequence 31, Appl
45	241.5	8.6	491	3	US-08-942-012B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-356-806-8
Sequence 8, Application US/09356806

Patent No. 6586175
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
TITLE OF INVENTION: 2B15 (UGT2B15) Genes
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
CURRENT FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-09-356-806-8

Query Match 93.0%; Score 2602.5; DB 4; Length 528;
Best Local Similarity 93.0%; Pred. No. 4.7e-263;
Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;

QY	1	MSKMTSALLLIOLSCYFSSGCGKYLWPTFESHMNNITTIIDELVORGHETVLAASA	60
DB	1	MSKMTSALLLIOLSCYFSSGCGKYLWPTFESHMNNITTIIDELVORGHETVLAASA	60
QY	61	SISDPGSPSTLKEFVYVSLTTEFEDIKOLVKRAELPKDTFMSYFQVQEIWTFN	120
DB	61	SISDPGSPSTLKEFVYVSLTTEFEDIKOLVKRAELPKDTFMSYFQVQEIWTFN	120
QY	121	DIILKPKDVISNKKMKLQESREPDVLAADVFPFELLAELIKIFVYSLSRSPGAI	180
DB	121	DIILKPKDVISNKKMKLQESREPDVLAADVFPFELLAELIKIFVYSLSRSPGAI	180
QY	181	EKHSGLLFPSPYVPMVSELSDMTIERVKNMIVLYFEFQIETDMKMDQFYSVL	240
DB	181	EKHSGLLFPSPYVPMVSELSDMTIERVKNMIVLYFEFQIETDMKMDQFYSVL	240
QY	241	GRPTTSLTAKADIMILIRNYWDQFPHPLIPNVEFGVGLCKCPAKPLPKMEEFVQSSG	300
DB	241	GRPTTSLTAKADIMILIRNYWDQFPHPLIPNVEFGVGLCKCPAKPLPKMEEFVQSSG	300
QY	301	ENGWVFSLSGMSVNTSEERANVIALAKIPOYLWRFQGNKPDITGLNRLTKWIPON	360
DB	301	ENGWVFSLSGMSVNTSEERANVIALAKIPOYLWRFQGNKPDITGLNRLTKWIPON	360

THIS PAGE BLANK (USP10)

QY	MSMKMTSALLILLOSCYFSSSGSGKVLWPTFSSHMNIKITLIDELVORGHEVTVLASSA	60
Db	1 MSVMXSVIILLIQLDSCFSSGNGCKVLVAALSYHMNIKITLIDELVORGHEVTVLASSA	60
QY	61 SISDPNSPSTLKEVYIPVSLYTEBEDITIKOLVYRMALPKDTWYSYFQVOEIMTFTN	120
Db	61 SILFDPNNSALLIEIYPTSLTKELENEIMQOIKRMSDLPDPTFWLTFYSQVOEIMSI	120
QY	121 DILRKFCKDVSNRKMKLKOESRGPVVLADAVPFGELMLKIPVYSGLRSPGVAL	180
Db	121 DILRKFCKDVSNRKMKLKOESRGPVVLADAVPFGELMLKIPVYSGLRSPGVAL	180
QY	181 EKSISGGLLPSPSYVPVVMSELSDOMTIERVXNMIVYLYFEFVFOIPFMKXMDPFYSEVL	240
Db	181 EKISGGFIFPSPSYVPVVMSELTQMTFEMERVXNMIVYLYFPDWFETIPFMKXMDPFYSEVL	240
QY	241 GRPTLTSETMAKADILIRNYDPFOPPHLLPNVEFVGLHCKKPKPLPKMEEPFVSSG	300
Db	241 GRPTLTSETMGKADWLIRNSMNFOPPYPLLPNVDFVGLHCKKPKPLPKMEEPFVSSG	300
QY	301 ENGCVVFSLSGMSVNTSEERANVIAALAKIPOKVLMRPDGGKPTTLGLNTRLKXMI	360
Db	301 ENGCVVFSLSGMSVNTSEERANVIAALAKIIOKVLMPRDGKPKPTTLGLNTRLKXMI	360
QY	361 DLIGHPTKAFITTHGGNGNGIYEAIYHGVPMVGVPIFGDOLNIAHMKAGAIVEINFKTM	420
Db	361 DLIGHPTKAFITTHGGNGNGIYEAIYHGVPMVGVPIFGDOLNIAHMKAGAIVEINFKTM	420
QY	421 TSIEDLPALETFVTDSGYENAMRSLRHDDPVPLBRAVWIEFVNRHNGAGHLSAA	480
Db	421 SSTDLNALAKRVINDNDSYEKNVKSRIHQDQVPLDPAFWLIEFVNRHNGAGHLSAA	480

RESULT 4

THIS PAGE BLANK (USRIU)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:49:45 ; Search time 36 Seconds

(without alignments)
2732.927 Million cell updates/sec

Title: US-09-980-729b-5

Perfect score: 2798

Sequence: 1 MSMKWTSAALLIOLSCYFSS.....KCLFSCCKFKTKRIEKRE 529

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 183983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

Result No.	Score	Query Match	Length	DB ID	Description
1	2602.5	93.0	528	15	US-10-205-522-8
2	2330	83.3	529	10	US-09-981-353-194
3	2330	83.3	529	15	US-10-057-834A-2
4	2322	83.0	524	15	US-10-205-522-40
5	2158.5	77.1	530	15	US-10-205-522-113
6	1924.5	68.8	527	10	US-09-981-353-166
7	1924.5	68.8	527	12	US-10-199-672-522
8	1924.5	68.8	527	12	US-10-187-749-522
9	1924.5	68.8	527	12	US-10-194-457-522
10	1924.5	68.8	527	12	US-10-184-642-522
11	1924.5	68.8	527	12	US-10-196-747-522
12	1924.5	68.8	527	12	US-10-173-689-522
13	1924.5	68.8	527	12	US-10-173-689-522
14	1924.5	68.8	527	12	US-10-173-691-522
15	1924.5	68.8	527	12	US-10-173-692-522

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

16	1924.5	68.8	527	12	US-10-173-694-522	Sequence 522, App
17	1924.5	68.8	527	12	US-10-173-698-522	Sequence 522, App
18	1924.5	68.8	527	12	US-10-173-699-522	Sequence 522, App
19	1924.5	68.8	527	12	US-10-173-707-522	Sequence 522, App
20	1924.5	68.8	527	12	US-10-174-569-522	Sequence 522, App
21	1924.5	68.8	527	12	US-10-174-583-522	Sequence 522, App
22	1924.5	68.8	527	12	US-10-174-587-522	Sequence 522, App
23	1924.5	68.8	527	12	US-10-174-589-522	Sequence 522, App
24	1924.5	68.8	527	12	US-10-175-591-522	Sequence 522, App
25	1924.5	68.8	527	12	US-10-175-736-522	Sequence 522, App
26	1924.5	68.8	527	12	US-10-175-742-522	Sequence 522, App
27	1924.5	68.8	527	12	US-10-175-744-522	Sequence 522, App
28	1924.5	68.8	527	12	US-10-175-745-522	Sequence 522, App
29	1924.5	68.8	527	12	US-10-175-751-522	Sequence 522, App
30	1924.5	68.8	527	12	US-10-175-751-522	Sequence 522, App
31	1924.5	68.8	527	12	US-10-175-754-522	Sequence 522, App
32	1924.5	68.8	527	12	US-10-176-480-522	Sequence 522, App
33	1924.5	68.8	527	12	US-10-176-489-522	Sequence 522, App
34	1924.5	68.8	527	12	US-10-176-754-522	Sequence 522, App
35	1924.5	68.8	527	12	US-10-176-755-522	Sequence 522, App
36	1924.5	68.8	527	12	US-10-176-759-522	Sequence 522, App
37	1924.5	68.8	527	12	US-10-176-920-522	Sequence 522, App
38	1924.5	68.8	527	12	US-10-176-922-522	Sequence 522, App
39	1924.5	68.8	527	12	US-10-176-924-522	Sequence 522, App
40	1924.5	68.8	527	12	US-10-176-984-522	Sequence 522, App
41	1924.5	68.8	527	12	US-10-179-508-522	Sequence 522, App
42	1924.5	68.8	527	12	US-10-179-512-522	Sequence 522, App
43	1924.5	68.8	527	12	US-10-179-515-522	Sequence 522, App
44	1924.5	68.8	527	12	US-10-173-702-522	Sequence 522, App
45	1924.5	68.8	527	12	US-10-173-703-522	Sequence 522, App

ALIGNMENTS

```
RESULT 1
US-10-205-522-8
Sequence: 8 Application US/10205522
Publication No. US20030077629A1
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/10/205,522
PRIOR FILING DATE: 2002-07-24
PRIOR APPLICATION NUMBER: US/09/356,806
PRIOR FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-10-205-522-8
Query Match 93.0%; Score 2602.5; DB 15; Length 528;
Best Local Similarity 93.0%; Pred. No. 8.5e-253; Indels 1; Gaps 1;
Matches 492; Conservative 14; Mismatches 22;
QY 1 MSMKWTSAALLIOLSCYFSSGCGKVLWPTFESHMNNIKITIDELVQRHEVTYLAASA 60
Db 1 MSMKWTSAALLIOLSCYFSSGCGKVLWPTFESHMNNIKITIDELVQRHEVTYLAASA 60
QY 61 SISFDPSBETLKFYVPSVLTTFEBDITKOLVKMAELPOTFMSYFQVOEIMWTFN 120
Db 61 SISFDPSBETLKFYVPSVLTTFEBDITKOLVKMAELPOTFMSYFQVOEIMWTFN 120
QY 121 DIRKFKDIVSKKLMKQLQESRFVDVLADAVPFGELLAELKLPFYVSLNFSFGYAI 180
```

THIS PAGE BLANK (USPTO)

```

Db 121 DILRFCKDIOVSNKKLMKKLOESRFDVVLADAVFPFGLLAEKLPVYSIRFSPGAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Qy 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300
Qy 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Db 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420
Db 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420
Qy 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDOVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Db 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDOVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Qy 481 HDLTFPOHYSIDVIGFLITCVATAIFLTKCFLFCQKFNKTRKIEKRE 529
Db 481 HDLTFPOHYSIDVIGFLITCVATAIFLTKCFLFCQKFNKTRKIEKRE 529

```

RESULT 2

```

US-09-981-353-194
; Sequence 194, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO: 194
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 088078CD1
US-09-981-353-194

```

Query Match

```

83.3%; Score 2330; DB 10; Length 529;
Best Local Similarity 81.1%; Pred. No. 2.3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```

```

Qy 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMANIKTILDELVORGEVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMANIKTILDELVORGEVTVLASSA 60
Qy 61 SISDPNPSPTLKFEVYPVSLTKTEPEDIIKOLVKMAELPKDTWMSYFSQVOEIMWTFN 120
Db 61 SISDPNPSPTLKFEVYPVSLTKTEPEDIIKOLVKMAELPKDTWMSYFSQVOEIMWTFN 120
Qy 121 DILRFCKDIOVSNKKLMKKLOESRFDVVLADAVFPFGLLAEKLPVYSIRFSPGAI 180
Db 121 DILRFCKDIOVSNKKLMKKLOESRFDVVLADAVFPFGLLAEKLPVYSIRFSPGAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Qy 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300

```

```

Qy 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Db 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420
Db 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420
Qy 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDOVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Db 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDOVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Qy 481 HDLTFPOHYSIDVIGFLITCVATAIFLTKCFLFCQKFNKTRKIEKRE 529
Db 481 HDLTFPOHYSIDVIGFLITCVATAIFLTKCFLFCQKFNKTRKIEKRE 529

```

RESULT 3

```

US-10-057-834A-2
; Sequence 2, Application US/10057834A
; Publication No. US2003009960A1
; GENERAL INFORMATION:
; APPLICANT: RATIN, MARK J.
; APPLICANT: INNOCENTI, FEDERICO
; APPLICANT: IYER, LALITHA
; APPLICANT: SAWYER, MICHAEL
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OPTIMIZING UGT2B7 SUBSTRATE DOSINGS
; FILE REFERENCE: ARCD:358US
; CURRENT APPLICATION NUMBER: US/10/057,834A
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: UNKNOWN
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-834A-2

```

```

Query Match 83.3%; Score 2330; DB 15; Length 529;
Best Local Similarity 81.1%; Pred. No. 2.3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```

```

Qy 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMANIKTILDELVORGEVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMANIKTILDELVORGEVTVLASSA 60
Qy 61 SISDPNPSPTLKFEVYPVSLTKTEPEDIIKOLVKMAELPKDTWMSYFSQVOEIMWTFN 120
Db 61 SISDPNPSPTLKFEVYPVSLTKTEPEDIIKOLVKMAELPKDTWMSYFSQVOEIMWTFN 120
Qy 121 DILRFCKDIOVSNKKLMKKLOESRFDVVLADAVFPFGLLAEKLPVYSIRFSPGAI 180
Db 121 DILRFCKDIOVSNKKLMKKLOESRFDVVLADAVFPFGLLAEKLPVYSIRFSPGAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYLFEPFOIPMKMKDQFSEVL 240
Qy 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNYWDFOPPHLLPNEVFVGLHCKPAKPLPKMEEFVQSSG 300
Qy 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Db 301 ENGUVVFSLSGMSVNTSEERANVIASALAKIPQKULMRFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420
Db 361 DLGHPKTKAFITTHGANGIYEAIYHGVPMVGPPIFGDQDNIAMKAKAAVEINFKTM 420

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:49:45 ; Search time 36 Seconds
(without alignments)
2732.927 Million cell updates/sec

Title: US-09-980-729b-5
Perfect score: 2798
Sequence: 1 MSKMTSALLLIQUSCYFSS.....KCFLFSCQKFKTKIEKRE 529

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues
Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2602.5	93.0	528	15	US-10-205-522-8
2	2330	83.3	529	10	US-09-981-353-194
3	2330	83.3	529	15	US-10-057-834A-2
4	2322	83.0	524	15	US-10-205-522-40
5	2158.5	77.1	530	15	US-10-205-522-113
6	1924.5	68.8	527	10	US-09-981-353-166
7	1924.5	68.8	527	12	US-10-199-672-522
8	1924.5	68.8	527	12	US-10-187-749-522
9	1924.5	68.8	527	12	US-10-194-457-522
10	1924.5	68.8	527	12	US-10-184-642-522
11	1924.5	68.8	527	12	US-10-196-747-522
12	1924.5	68.8	527	12	US-10-173-689-522
13	1924.5	68.8	527	12	US-10-173-690-522
14	1924.5	68.8	527	12	US-10-173-691-522
15	1924.5	68.8	527	12	US-10-173-692-522

16	1924.5	68.8	527	12	US-10-173-694-522	Sequence 522, App
17	1924.5	68.8	527	12	US-10-173-698-522	Sequence 522, App
18	1924.5	68.8	527	12	US-10-173-699-522	Sequence 522, App
19	1924.5	68.8	527	12	US-10-173-707-522	Sequence 522, App
20	1924.5	68.8	527	12	US-10-174-569-522	Sequence 522, App
21	1924.5	68.8	527	12	US-10-174-583-522	Sequence 522, App
22	1924.5	68.8	527	12	US-10-174-587-522	Sequence 522, App
23	1924.5	68.8	527	12	US-10-174-589-522	Sequence 522, App
24	1924.5	68.8	527	12	US-10-174-591-522	Sequence 522, App
25	1924.5	68.8	527	12	US-10-175-736-522	Sequence 522, App
26	1924.5	68.8	527	12	US-10-175-742-522	Sequence 522, App
27	1924.5	68.8	527	12	US-10-175-745-522	Sequence 522, App
28	1924.5	68.8	527	12	US-10-175-745-522	Sequence 522, App
29	1924.5	68.8	527	12	US-10-175-748-522	Sequence 522, App
30	1924.5	68.8	527	12	US-10-175-751-522	Sequence 522, App
31	1924.5	68.8	527	12	US-10-175-754-522	Sequence 522, App
32	1924.5	68.8	527	12	US-10-176-480-522	Sequence 522, App
33	1924.5	68.8	527	12	US-10-176-489-522	Sequence 522, App
34	1924.5	68.8	527	12	US-10-176-754-522	Sequence 522, App
35	1924.5	68.8	527	12	US-10-176-755-522	Sequence 522, App
36	1924.5	68.8	527	12	US-10-176-759-522	Sequence 522, App
37	1924.5	68.8	527	12	US-10-176-920-522	Sequence 522, App
38	1924.5	68.8	527	12	US-10-176-922-522	Sequence 522, App
39	1924.5	68.8	527	12	US-10-176-924-522	Sequence 522, App
40	1924.5	68.8	527	12	US-10-176-984-522	Sequence 522, App
41	1924.5	68.8	527	12	US-10-179-508-522	Sequence 522, App
42	1924.5	68.8	527	12	US-10-179-512-522	Sequence 522, App
43	1924.5	68.8	527	12	US-10-179-515-522	Sequence 522, App
44	1924.5	68.8	527	12	US-10-173-702-522	Sequence 522, App
45	1924.5	68.8	527	12	US-10-173-703-522	Sequence 522, App

ALIGNMENTS

RESULT 1
US-10-205-522-8
Sequence-8: Application US/10205522
Publication No. US20030077629A1
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT FILING DATE: US/10/205,522
PRIORITY FILING DATE: 2002-07-24
PRIORITY FILING DATE: US/09/356,806
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-10-205-522-8
Query Match 93.0%; Score 2602.5; DB 15; Length 528;
Best Local Similarity 93.0%; Pred. No. 8.5e-253; Indels 1; Gaps 1;
Matches 492; Conservative 14; Mismatches 22;
QY 1 MSKMTSALLLIQUSCYFSSGCGKVLWPTFSHMNNTITIDELVQHGVEVTLASSA 60
Db 1 MSKMTSALLLIQUSCYFSSGCGKVLWPTFSHMNNTITIDELVQHGVEVTLASSA 60
QY 61 SISFDPSBETLKFVYPSVLTTEPEDITKOLVKRAELPKOTFMSYFQVQEIWMFTFN 120
Db 61 SISFDPSBETLKFVYPSVLTTEPEDITKOLVKRAELPKOTFMSYFQVQEIWMFTFN 120
QY 121 DIRKFKDIVSKKMKLQESRFDVLAIVFPFELLAEELKIPFYVSLRFSPEYAI 180

Db 121 DLRKFCDDIVSNKKLMKKLOESRFDVVLADAVPFGEILLAEILKI PFVYSLRSPGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Db 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Db 421 SSTDLNALKRVINDPSYKENAMRLSRIHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Qy 481 HDLTFQYHSIDVIGFLLTCVATAIPLFTKCFPLSCQKFNKTRKIERE 529
Db 481 HDLTFQYHSIDVIGFLLTCVATAIPLFTKCFPLSCQKFNKTRKIERE 529

RESULT 2
US-09-981-353-194

; Sequence 194, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 194
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 088078CD1
US-09-981-353-194

Query Match 83.3%; Score 2330; DB 10; Length 529;
Best Local Similarity 81.1%; Pred. No. 2,3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;
Qy 1 MSKWTSAALLIIQISCFYSSGSGKVLVWPTFSHMNIKITIDELVQGHVTVLASSA 60
Db 1 MSVKMTSVILLIIQISCFSSGNGKVLVMAEYSHMNIKITIDELIQRGHEVTVLASSA 60
Qy 61 SISFDPSPTLKEVYVVSILTKTEFEDIKQLYKRAELPKDFTWVSFQVOEIMMTFN 120
Db 61 SILFDPNNSALKIEIYPTSLTKTELENFIQOIKRMSDLPKDTFWLVSQVOEIMSYFG 120
Qy 121 DLRKFCDDIVSNKKLMKKLOESRFDVVLADAVPFGEILLAEILKI PFVYSLRSPGAI 180
Db 121 DLRKFCDDIVSNKKLMKKLOESRFDVVLADAVPFGEILLAEILKI PFVYSLRSPGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300

Qy 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Db 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Db 421 SSTDLNALKRVINDPSYKENAMRLSRIHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Qy 481 HDLTFQYHSIDVIGFLLTCVATAIPLFTKCFPLSCQKFNKTRKIERE 529
Db 481 HDLTFQYHSIDVIGFLLTCVATAIPLFTKCFPLSCQKFNKTRKIERE 529

RESULT 3
US-10-057-834A-2

; Sequence 2, Application US/10057834A
; Publication No. US20030099960A1
; GENERAL INFORMATION:
; APPLICANT: RATAIN, MARK J.
; APPLICANT: INNOCENTI, FEDERICO
; APPLICANT: DAS, SOMA
; APPLICANT: IYER, LALITHA
; APPLICANT: SAWYER, MICHAEL
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OPTIMIZING UGT2B7 SUBSTRATE TOXICITY
; FILE REFERENCE: ARCD:35805
; CURRENT APPLICATION NUMBER: US/10/057,834A
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: UNKNOWN
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-834A-2

Query Match 83.3%; Score 2330; DB 15; Length 529;
Best Local Similarity 81.1%; Pred. No. 2,3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;
Qy 1 MSKWTSAALLIIQISCFYSSGSGKVLVWPTFSHMNIKITIDELVQGHVTVLASSA 60
Db 1 MSVKMTSVILLIIQISCFSSGNGKVLVMAEYSHMNIKITIDELIQRGHEVTVLASSA 60
Qy 61 SISFDPSPTLKEVYVVSILTKTEFEDIKQLYKRAELPKDFTWVSFQVOEIMMTFN 120
Db 61 SILFDPNNSALKIEIYPTSLTKTELENFIQOIKRMSDLPKDTFWLVSQVOEIMSYFG 120
Qy 121 DLRKFCDDIVSNKKLMKKLOESRFDVVLADAVPFGEILLAEILKI PFVYSLRSPGAI 180
Db 121 DLRKFCDDIVSNKKLMKKLOESRFDVVLADAVPFGEILLAEILKI PFVYSLRSPGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFWMFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Db 301 ENGIVVPSLGSVMSVNTSEERANVIASALAKIPOKVLRFPGDKNKPDTGLNTRLYKWIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPILFADQPDNIAMKAKGAVALINFTM 420

Db 361 DLGHPKTRAFITGANGIYEALYHGI PMVGIPLEFADQPDNIAMKARGAIVDRNTM 420
Qy 421 TSEDLRLARLTVITDSSYKENAMRLSRHHDQPKPLDRAVFWIEFVNRHKGALHLSAA 480
Db 421 SSTDLNALKRVINDPSYKENAMKLSRIQHDPKPLDRAVFWIEFVNRHKGALHLSAA 480
Qy 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRKIEKRE 529
Db 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRKIEKRE 529

RESULT 4
US-10-205-522-40
; Sequence 40, Application US/10205522
; Publication No. US20030077629A1
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/10/205,522
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US/09/356,806
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 524
; TYPE: PRF
; ORGANISM: H. sapiens
US-10-205-522-40

Query Match 83.0%; Score 2322; DB 15; Length 524;
Best Local Similarity 81.5%; Pred. No. 1.4e-224;
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0;

Qy 1 MSMKWTSAALLLIQUSCFSSGSGKVLVWPTFESHMNNITITDELVORGHETVLAASA 60
Db 1 MSVMTSVILLIIQUSCFSSGSGKVLVWMAEYSHMNNIKITIDELLORGHETVLAASA 60
Qy 61 SISDRPSPSTLKEVPVSLTKTEPEDIIKQLYKRAELPKDTFMSYFSQVOEIMTFTN 120
Db 61 SILDPNNSSALKIEIYPTSLTKTELEBNFIQOIKRMSDLPKDTFMSYFSQVOEIMSI 120
Qy 121 DILKPKCDIYSNKKMLKLOESRFDVVLADAVFPFGELLAELKIPFVYSLRSPGYAI 180
Db 121 DITKPKCDVSNKKPKKQVQESRFDVIFADALFPCELLAELNIPFVYSLRSPGYTF 180
Qy 181 EKHSGLLFPSPYVPMVMSLSDQMTFIERVKMIIYLYFEFMOIFDMKKMDQFYSEVL 240
Db 181 EKHSGLLFPSPYVPMVMSLSDQMTFIERVKMIIYLYFEFMOIFDMKKMDQFYSEVL 240
Qy 241 GRPPTLSEYMAKADIWLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPPTLSEYMAKADIWLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGUVVSLGSMVNTSEERANVIALAKIPQKVLMPFGNKPDTGLNTRLYKMI 360
Db 301 ENGUVVSLGSMVNTSEERANVIALAKIPQKVLMPFGNKPDTGLNTRLYKMI 360
Qy 361 DLGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQPDNIAMKARGAIVDRNTM 420
Db 361 DLGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQPDNIAMKARGAIVDRNTM 420
Qy 421 TSEDLRLARLTVITDSSYKENAMRLSRHHDQPKPLDRAVFWIEFVNRHKGALHLSAA 480
Db 421 SSTDLNALKRVINDPSYKENAMKLSRIQHDPKPLDRAVFWIEFVNRHKGALHLSAA 480
Qy 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRK 524

Db 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRKAK 524

RESULT 5
US-10-205-522-113
; Sequence 113, Application US/10205522
; Publication No. US20030077629A1
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/10/205,522
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US/09/356,806
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 530
; TYPE: PRF
; ORGANISM: H. sapiens
US-10-205-522-113

Query Match 77.1%; Score 2158.5; DB 15; Length 530;
Best Local Similarity 74.5%; Pred. No. 4.2e-208;
Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

Qy 1 MSMKWTSAALLLIQUSCFSSGSGKVLVWPTFESHMNNITITDELVORGHETVLAASA 60
Db 1 MSKMTSVILLIIQUSCFSSGSGKVLVWPTFESHMNNIKITIDELLORGHETVLAASA 60
Qy 61 SISDRPSPSTLKEVPVSLTKTEPEDIIKQLYKRAELPKDTFMSYFSQVOEIMTFTN 119
Db 61 SILDPNNSSALKIEIYPTSLTKTELEBNFIQOIKRMSDLPKDTFMSYFSQVOEIMSI 120
Qy 120 NDILKPKCDIYSNKKMLKLOESRFDVVLADAVFPFGELLAELKIPFVYSLRSPGYA 179
Db 120 YDYSNKKLCKRAVYLNKKMLKLOESRFDVVLADALNPGELLAELNIPFVYSLRSPGYT 180
Qy 180 IEHSGGLLFPSPYVPMVMSLSDQMTFIERVKMIIYLYFEFMOIFDMKKMDQFYSEV 239
Db 180 IEHSGGLLFPSPYVPMVMSLSDQMTFIERVKMIIYLYFEFMOIFDMKKMDQFYSEV 240
Qy 240 LGRPTLSEYMAKADIWLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSS 299
Db 240 LGRPTLSEYMAKADIWLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSS 300
Qy 300 ENGUVVSLGSMVNTSEERANVIALAKIPQKVLMPFGNKPDTGLNTRLYKMI 359
Db 300 GENGIVVSLGSMVNTSEERANVIALAKIPQKVLMPFGNKPDTGLNTRLYKMI 360
Qy 360 NDILGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQPDNIAMKARGAIVDRNTM 419
Db 360 NDILGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQPDNIAMKARGAIVDRNTM 420
Qy 420 MTESEDLRLARLTVITDSSYKENAMRLSRHHDQPKPLDRAVFWIEFVNRHKGALHLSAA 479
Db 420 MTESEDLRLARLTVITDSSYKENAMRLSRHHDQPKPLDRAVFWIEFVNRHKGALHLSAA 480
Qy 480 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRKIEKRE 529
Db 480 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCOKFNKTRKIEKRE 529

Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Laeek, Amy W.
APPLICANT: Jones, David A.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 166
LENGTH: 527
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 2434655CD1
US-09-981-353-166

Query Match 68.8%; Score 1924.5; DB 10; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184; Indels 3; Gaps 2;
Matches 369; Conservative 47; Mismatches 109;

QY 2 SMKMTSALLLIQLSCYFSSGCGKLVLPTEFSSHMNIKITLDELVORGEVTVLASAS 61
DB 3 SDKSALVFLLLQLFC-VGCGFCGKLVPCDMSHMLNVKVLLEBLIYRGHEVTVLTHSKP 61
QY 62 ISFPDNGSTLKEFVYVSLTKTEFEDIIKQLVKRMALPKDTFMSYQOELMTFND 121
DB 62 SLIDYRKPSALKFEVVMHPDRTENEI FVDLALN--VLPELSTWQSVIKLNDFFVEIRG 119
QY 122 ILRKFCXDIVSNKKLMKKLOESRFDVVLADAVPEFGBELAEELKIPVYSIRFSPGVAIE 181
DB 120 TLKMMCESFTYNQTLMMKKLOETNYDVMLIDPVICGDLMAELAVPVLTLRISVGME 179
QY 182 KHSGLLFPSPYVPMVSELSQDQMTFIERVKMIVLYFEFVFOIFDMKKADQYSEVYG 241
DB 180 RSCGKLPAPLSYVPMVGLDRMTFLERVKNMSLSVLFHFMIQDYDHFMEEFYSKALG 239
QY 242 RPTTSETMAKADIVLIRNYWDFOPPHLLPNVEFVGLHCKPAKPLPKEMEFPVQSSGE 301
DB 240 RPTTLCETVGAKEIWLIRTYWDFEFPPOYPFEEFVGLHCKPAKPLPKEMENFVQSSGE 299
QY 302 NGVVVFSLSGWSNTESEERANVIALAKIPQKVLMPFDGNKPDTLGLNTRLYKMIPOND 361
DB 300 DGIYVFSLSGLFQVNTSEKANIISALAQIPQKVLMPFKGKPSSTLGANTRLYDMPOND 359
QY 362 LLGHPKTAFTTHGGMNGIYEAIIYHGVPMGVPIFGDQLDNIAMKAKGAAVEINFKTM 421
DB 360 LLGHPKTAFTTHGGMNGIYEAIIYHGVPMGVPIFGDQLDNIAMKAKGAAVEINFKTM 419
QY 422 SEDLLRALRTVITDSSYKENAMRLSRIRHDDPVKPLDRAVWIEFVMMHKKAKHLRSAAH 481
DB 420 SEDLLRALRTVITDSSYKENAMRLSRIRHDDPVKPLDRAVWIEFVMMHKKAKHLRSAAH 479
QY 482 DLTWFOHYSIDVIGFLTCTVATAIFLFTKCFPFSCQKFNKTRKIEKRE 529
DB 480 DLTWFOHYSIDVIGFLTCTVATAIFLFTKCFPFSCQKFNKTRKIEKRE 527

RESULT 7
US-10-199-672-522
Sequence 522, Application US/10199672
Publication No. US2003014842A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/199,672
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-199-672-522

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKMTSALLLIQLSCYSSGCGKLVLPTEFSSHMNIKITLDELVORGEVTVLASAS 61
DB 3 SDKSALVFLLLQLFC-VGCGFCGKLVLPDMSHMLNVKVLLEBLIYRGHEVTVLTHSKP 61
QY 62 ISFPDNGSTLKEFVYVSLTKTEFEDIIKQLVKRMALPKDTFMSYQOELMTFND 121
DB 62 SLIDYRKPSALKFEVVMHPDRTENEI FVDLALN--VLPELSTWQSVIKLNDFFVEIRG 119
QY 122 ILRKFCXDIVSNKKLMKKLOESRFDVVLADAVPEFGBELAEELKIPVYSIRFSPGVAIE 181
DB 120 TLKMMCESFTYNQTLMMKKLOETNYDVMLIDPVICGDLMAELAVPVLTLRISVGME 179
QY 182 KHSGLLFPSPYVPMVSELSQDQMTFIERVKMIVLYFEFVFOIFDMKKADQYSEVYG 241
DB 180 RSCGKLPAPLSYVPMVGLDRMTFLERVKNMSLSVLFHFMIQDYDHFMEEFYSKALG 239
QY 242 RPTTSETMAKADIVLIRNYWDFOPPHLLPNVEFVGLHCKPAKPLPKEMEFPVQSSGE 301
DB 240 RPTTLCETVGAKEIWLIRTYWDFEFPPOYPFEEFVGLHCKPAKPLPKEMENFVQSSGE 299
QY 302 NGVVVFSLSGWSNTESEERANVIALAKIPQKVLMPFDGNKPDTLGLNTRLYKMIPOND 361
DB 300 DGIYVFSLSGLFQVNTSEKANIISALAQIPQKVLMPFKGKPSSTLGANTRLYDMPOND 359
QY 362 LLGHPKTAFTTHGGMNGIYEAIIYHGVPMGVPIFGDQLDNIAMKAKGAAVEINFKTM 421
DB 360 LLGHPKTAFTTHGGMNGIYEAIIYHGVPMGVPIFGDQLDNIAMKAKGAAVEINFKTM 419
QY 422 SEDLLRALRTVITDSSYKENAMRLSRIRHDDPVKPLDRAVWIEFVMMHKKAKHLRSAAH 481
DB 420 SEDLLRALRTVITDSSYKENAMRLSRIRHDDPVKPLDRAVWIEFVMMHKKAKHLRSAAH 479
QY 482 DLTWFOHYSIDVIGFLTCTVATAIFLFTKCFPFSCQKFNKTRKIEKRE 529


```

QY 2 SMKWTALLIOLSCYFSSGCGKVLWPTFESHMNNIKTILDELVORGHEVTVLASSAS 61
DB 3 SDSKALVFLLLQLFC-VGCGFCGKVLWPCMSHMLNVKVLLEELIVRGHEVTVLTHSKP 61
QY 62 ISFDPNBSPSTLKFVYVPSLTKTTEFEDIKOLVWRMAELPKDTFWSYFSQVOEIMTFND 121
DB 62 SLIDYRKPALKFEVYVHPQDRTEENELFVDLALN--VLPGLSWTQSVITKLNDFVEIRG 119
QY 122 ILKFKCDIVSNKKMKLQESRFDVLAADVPPGELLABELLKIIPVYSLRFSGYAIE 181
DB 120 TKMKCCSFYINQTLMKKLOETNYDVMLIDPVI PCGDLMAELLAVPVLTLRISVGME 179
QY 182 KHSGGLFPSPSYVWVMSLSDQMTFIERVKNMIVLYFEFPMQIPDMKKMDQFYSVLG 241
DB 180 RSCGKLPAPLSYVVPMTGLTDRMTFLERVKSNMISVLFHFWDYDHFMEEFYSKALG 239
QY 242 RPTLSETMAKADWLIRNYWDFOPPHLBNVEFVGLCKCPAKPLPKEMEDEFVSSGE 301
DB 240 RPTLCEVGAELWILRTYWDPEFPYOPNPEFVGLCKCPAKALPKEMENFVSSGE 299
QY 302 NGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPTDGLNTRLYKWIPOND 361
DB 300 DGIWVFSLSGLFQVWTEERANIISALAQIPQKVLMPYKGGKSTLGANTRLYDWMIPOND 359
QY 362 LKGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDOLNIAHMKAKGAAVEINFKMT 421
DB 360 LKGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDOLNIAHMKAKGAAVEINFKMT 419
QY 422 SEDLRALRTYITDSSYKENAMRLSRINHDPVKPLDRAVFWIEFVRHKGAKHLRSAAH 481
DB 420 SEDLRALRTYITDSSYKENAMRLSRINHDPVKPLDRAVFWIEFVRHKGAKHLRSAAH 479
QY 482 DLTWFOHYSIDVIGFLTVCVATAIFLFTKCFLFSQCFKFNKTRKIEKRE 529
DB 480 DLTWFOHYSIDVIGFLTVCVATAIFLFTKCFLFSQCFKFNKTRKIEKRE 527

```

RESULT 10
US-10-184-642-522
; Sequence 522, Application US/10184642
; Publication No. US20030157635A1

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jilan
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C194
; CURRENT FILING DATE: 2002-06-27
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-642-522

```

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKWTALLIOLSCYFSSGCGKVLWPTFESHMNNIKTILDELVORGHEVTVLASSAS 61
DB 3 SDSKALVFLLLQLFC-VGCGFCGKVLWPCMSHMLNVKVLLEELIVRGHEVTVLTHSKP 61

```

DB 3 SDSKALVFLLLQLFC-VGCGFCGKVLWPCMSHMLNVKVLLEELIVRGHEVTVLTHSKP 61
QY 62 ISFDPNBSPSTLKFVYVPSLTKTTEFEDIKOLVWRMAELPKDTFWSYFSQVOEIMTFND 121
DB 62 SLIDYRKPALKFEVYVHPQDRTEENELFVDLALN--VLPGLSWTQSVITKLNDFVEIRG 119
QY 122 ILKFKCDIVSNKKMKLQESRFDVLAADVPPGELLABELLKIIPVYSLRFSGYAIE 181
DB 120 TKMKCCSFYINQTLMKKLOETNYDVMLIDPVI PCGDLMAELLAVPVLTLRISVGME 179
QY 182 KHSGGLFPSPSYVWVMSLSDQMTFIERVKNMIVLYFEFPMQIPDMKKMDQFYSVLG 241
DB 180 RSCGKLPAPLSYVVPMTGLTDRMTFLERVKSNMISVLFHFWDYDHFMEEFYSKALG 239
QY 242 RPTLSETMAKADWLIRNYWDFOPPHLBNVEFVGLCKCPAKPLPKEMEDEFVSSGE 301
DB 240 RPTLCEVGAELWILRTYWDPEFPYOPNPEFVGLCKCPAKALPKEMENFVSSGE 299
QY 302 NGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPTDGLNTRLYKWIPOND 361
DB 300 DGIWVFSLSGLFQVWTEERANIISALAQIPQKVLMPYKGGKSTLGANTRLYDWMIPOND 359
QY 362 LKGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDOLNIAHMKAKGAAVEINFKMT 421
DB 360 LKGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDOLNIAHMKAKGAAVEINFKMT 419
QY 422 SEDLRALRTYITDSSYKENAMRLSRINHDPVKPLDRAVFWIEFVRHKGAKHLRSAAH 481
DB 420 SEDLRALRTYITDSSYKENAMRLSRINHDPVKPLDRAVFWIEFVRHKGAKHLRSAAH 479
QY 482 DLTWFOHYSIDVIGFLTVCVATAIFLFTKCFLFSQCFKFNKTRKIEKRE 529
DB 480 DLTWFOHYSIDVIGFLTVCVATAIFLFTKCFLFSQCFKFNKTRKIEKRE 527

```

RESULT 11
US-10-196-747-522
; Sequence 522, Application US/10196747
; Publication No. US20030162250A1

GENERAL INFORMATION:

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jilan
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C346
; CURRENT FILING DATE: 2002-07-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-196-747-522

```

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKWTALLIOLSCYFSSGCGKVLWPTFESHMNNIKTILDELVORGHEVTVLASSAS 61
DB 3 SDSKALVFLLLQLFC-VGCGFCGKVLWPCMSHMLNVKVLLEELIVRGHEVTVLTHSKP 61
QY 62 ISFDPNBSPSTLKFVYVPSLTKTTEFEDIKOLVWRMAELPKDTFWSYFSQVOEIMTFND 121

```
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
Qy 122 ILRKFCDIVSNKKMLKQESRFDVVLADAVFPFGELLAELKIPFVYSLRSPGYAIE 181
Qy 120 TLKMMCESFIYNTQTLMKKQOETNYDVLIDPVI PCGDLMAELLAVPVLTLRLISVGNGME 179
Db 182 KHSGLLFPSPSYVPMVMSLSDQMTFIERVKNMIVLYFEFPOIFPMKKMDQFSEVLG 241
Db 180 RSCGKLPAPIASYVPVPMPTGLTDRMTFLERVKNSMLSTLFHFWMIDYIHWEERYSALG 239
Qy 242 RPTLTSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEERVOSSGE 301
Db 240 RPTLTCEVGAKEIWLIRTYWDFEPQYQPNFEPVGLHCKPAKALPKEMENVOSSGE 299
Qy 302 NGVVFSLGSMVNTSEERANVIASALAKIPQKYLWRFDNKPDGLNTRLYKWIQND 361
Db 300 DGIIVFSLGSLFQWVTEKANIIASALAQIPQKYLWRYKCKRSTLGANTRLYDWMIPND 359
Qy 362 LIGHPKTAKFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAAVEINFKTMT 421
Db 360 LIGHPKTAKFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAAVEINFKTMT 419
Qy 422 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 481
Db 420 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 479
Qy 482 DLTWFOHYSIDVIGFLLTCVATAIFLTCKFLSCQKFNTRKIEKRE 529
Db 480 DLTWFOHYSIDVIGFLLTCVATAIFLTCKFLSCQKFNTRKIEKRE 527
```

RESULT 12

```
US-10-173-689-522
; Sequence 522, Application US/10173689
; Publication No. US20030166104A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C10
; CURRENT APPLICATION NUMBER: US/10/173,689
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-689-522
```

```
Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
Qy 2 SMKWTSAALLLIQSCYFSSGCGKVLWPTFESHMMNITLIDELVORGHEVTLASSAS 61
Db 3 SDRKALVFLILQIFC-VGCGFCGKVLWPCDMSHMLNVKILBELIVRGHEVTLASSAS 61
Qy 62 ISFDPNSPSTLKEFVPMVSLTKTEFEDIIKQLVKNMAELPKDTFWSYFSQVOEIMTFFND 121
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
```

```
Qy 122 ILRKFCDIVSNKKMLKQESRFDVVLADAVFPFGELLAELKIPFVYSLRSPGYAIE 181
Db 120 TLKMMCESFIYNTQTLMKKQOETNYDVLIDPVI PCGDLMAELLAVPVLTLRLISVGNGME 179
Qy 182 KHSGLLFPSPSYVPMVMSLSDQMTFIERVKNMIVLYFEFPOIFPMKKMDQFSEVLG 241
Db 180 RSCGKLPAPIASYVPVPMPTGLTDRMTFLERVKNSMLSTLFHFWMIDYIHWEERYSALG 239
Qy 242 RPTLTSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEERVOSSGE 301
Db 240 RPTLTCEVGAKEIWLIRTYWDFEPQYQPNFEPVGLHCKPAKALPKEMENVOSSGE 299
Qy 302 NGVVFSLGSMVNTSEERANVIASALAKIPQKYLWRFDNKPDGLNTRLYKWIQND 361
Db 300 DGIIVFSLGSLFQWVTEKANIIASALAQIPQKYLWRYKCKRSTLGANTRLYDWMIPND 359
Qy 362 LIGHPKTAKFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAAVEINFKTMT 421
Db 360 LIGHPKTAKFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAAVEINFKTMT 419
Qy 422 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 481
Db 420 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 479
Qy 482 DLTWFOHYSIDVIGFLLTCVATAIFLTCKFLSCQKFNTRKIEKRE 529
Db 480 DLTWFOHYSIDVIGFLLTCVATAIFLTCKFLSCQKFNTRKIEKRE 527
```

RESULT 13

```
US-10-173-690-522
; Sequence 522, Application US/10173690
; Publication No. US20030166105A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C9
; CURRENT APPLICATION NUMBER: US/10/173,690
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-690-522
```

```
Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
Qy 2 SMKWTSAALLLIQSCYFSSGCGKVLWPTFESHMMNITLIDELVORGHEVTLASSAS 61
Db 3 SDRKALVFLILQIFC-VGCGFCGKVLWPCDMSHMLNVKILBELIVRGHEVTLASSAS 61
Qy 62 ISFDPNSPSTLKEFVPMVSLTKTEFEDIIKQLVKNMAELPKDTFWSYFSQVOEIMTFFND 121
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
Qy 122 ILRKFCDIVSNKKMLKQESRFDVVLADAVFPFGELLAELKIPFVYSLRSPGYAIE 181
Db 120 TLKMMCESFIYNTQTLMKKQOETNYDVLIDPVI PCGDLMAELLAVPVLTLRLISVGNGME 179
```

182 KHSGGLPPSPVYVPMSELSDQMTFIERVKNMIVLYFEFVFOI.FDKMKDOFYSEVIG 241
180 RSCGKLPAPLSYVPMPTGLTRMTFLERVKNMISVLFHFMIQDYDHFHEEFSKALG 239
242 RPTLSETMAKADIWLIRNWDFOFPHPLLPNVEFVGLHCKPAKPLPKEMEERFVQSSGE 301
240 RPTLCEVGAKEIWLIRTYWDFEFPQYQPNFEFVGLHCKPAKPLPKEMENFVQSSGE 299
302 NGVVVFSLSGMSVNSTSERANVIAALAKIPQKULMRDGNKRPDLGLNTRLYKMIPOND 361
300 DGIIVFSLGSLFQVWTEBKANIISALAQIPQKULMRKGGKPSLTGANTLRYDMPOND 359
362 LIGHPKTAFITTHGNGMIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAIVEINFKMT 421
360 LIGHPKTAFITTHGNGMIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAIVEINFKMT 419
422 SEDLLRALRTVITDSSYKENAMRLSRTHHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAH 481
420 SEDLLRALRTVITDSSYKENAMRLSRTHHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAH 479
482 DLTWFOHYSIDVIGFLITCVATAIFLFTKCFLFSQCKENKTRKIEKRE 529
480 DLTWFOHYSIDVIGFLITCVATAIFLFTKCFLFSQCKENKTRKIEKRE 527

RESULT 14

US-10-173-691-522
Sequence 522, Application US/10173691
Publication No. US20030166106A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jlan
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C13
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-173-691-522

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
2 SMKTSALLIQLSCYFSSGCGKVLWMPTEFSHMNIKITIDELVORGHEVTVLASSAS 61
3 SDKSALVFLIQLFC-VGCGFCGKVLWPCDMSHMLNKVILEELIVGHEVTVLTHSKP 61
62 ISFPNSSTLKFVYVPSLTKEFEDIKQIVKMAELPKDTFWSYFSQVOEIMWTEND 121
62 SLIDYRKPSALKFEVVMHPODRTENEI.FVDLALN--VLPGISTWQSYIKLNDPFVEIRG 119
122 ILRKCCKOIVSNKKMLKQESRPDVVLADAVFPFGEILAEELKIPYYSLRFSBGAYIE 181
120 TLKMKCESFIYQTLMKKLOETNYDVMLIDPVI.PCGDLAELBLAVPVLTLRISVGME 179
182 KHSGGLPPSPVYVPMSELSDQMTFIERVKNMIVLYFEFVFOI.FDKMKDOFYSEVIG 241

180 RSCGKLPAPLSYVPMPTGLTRMTFLERVKNMISVLFHFMIQDYDHFHEEFSKALG 239
242 RPTLSETMAKADIWLIRNWDFOFPHPLLPNVEFVGLHCKPAKPLPKEMEERFVQSSGE 301
240 RPTLCEVGAKEIWLIRTYWDFEFPQYQPNFEFVGLHCKPAKPLPKEMENFVQSSGE 299
302 NGVVVFSLSGMSVNSTSERANVIAALAKIPQKULMRDGNKRPDLGLNTRLYKMIPOND 361
300 DGIIVFSLGSLFQVWTEBKANIISALAQIPQKULMRKGGKPSLTGANTLRYDMPOND 359
362 LIGHPKTAFITTHGNGMIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAIVEINFKMT 421
360 LIGHPKTAFITTHGNGMIYEAIYHGVPMGVPIFGQDLNIAHMKAKGAIVEINFKMT 419
422 SEDLLRALRTVITDSSYKENAMRLSRTHHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAH 481
420 SEDLLRALRTVITDSSYKENAMRLSRTHHDOVPKPLDRAVFWIEFVNRHKGAKHLRSAH 479
482 DLTWFOHYSIDVIGFLITCVATAIFLFTKCFLFSQCKENKTRKIEKRE 529
480 DLTWFOHYSIDVIGFLITCVATAIFLFTKCFLFSQCKENKTRKIEKRE 527

RESULT 15

US-10-173-692-522
Sequence 522, Application US/10173692
Publication No. US20030166188A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jlan
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C20
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-173-692-522

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
2 SMKTSALLIQLSCYFSSGCGKVLWMPTEFSHMNIKITIDELVORGHEVTVLASSAS 61
3 SDKSALVFLIQLFC-VGCGFCGKVLWPCDMSHMLNKVILEELIVGHEVTVLTHSKP 61
62 ISFPNSSTLKFVYVPSLTKEFEDIKQIVKMAELPKDTFWSYFSQVOEIMWTEND 121
62 SLIDYRKPSALKFEVVMHPODRTENEI.FVDLALN--VLPGISTWQSYIKLNDPFVEIRG 119
122 ILRKCCKOIVSNKKMLKQESRPDVVLADAVFPFGEILAEELKIPYYSLRFSBGAYIE 181
120 TLKMKCESFIYQTLMKKLOETNYDVMLIDPVI.PCGDLAELBLAVPVLTLRISVGME 179
182 KHSGGLPPSPVYVPMSELSDQMTFIERVKNMIVLYFEFVFOI.FDKMKDOFYSEVIG 241
180 RSCGKLPAPLSYVPMPTGLTRMTFLERVKNMISVLFHFMIQDYDHFHEEFSKALG 239
242 RPTLSETMAKADIWLIRNWDFOFPHPLLPNVEFVGLHCKPAKPLPKEMEERFVQSSGE 301

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:24 ; Search time 47 Seconds
(without alignments)
1786.518 Million cell updates/sec

Title: US-09-980-729B-5
2798
Perfect score: 1 MSKMFTSALLIQLSCYFSS.....KCFLFSCQKFNKTRIEKRE 529
Sequence:

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2798	100.0	529	21	AA28677
2	2602.5	93.0	528	21	AA78933
3	2584.5	92.4	540	22	ABG05525
4	2503	89.5	529	23	AA022794
5	2330	83.3	529	23	AA28617
6	2325	83.1	533	22	ABG05523
7	2322	83.0	524	21	AA78934
8	2292	81.9	529	22	AA202188
9	2233	79.8	532	22	ABG05524

10	2169.5	77.5	530	19	AAW47126
11	2158.5	77.1	530	21	AA78933
12	2158.5	77.1	530	24	ABJ19806
13	1924.5	68.8	527	22	AAU29284
14	1924.5	68.8	527	23	AAU77927
15	1924.5	68.8	527	23	AAE15434
16	1924.5	68.8	527	24	AAU54567
17	1924.5	68.8	527	24	ABU71372
18	1924.5	68.8	527	24	ABU65829
19	1924.5	68.8	527	24	ABU61652
20	1924.5	68.8	527	24	ABU67666
21	1924.5	68.8	527	24	ABU65524
22	1924.5	68.8	527	24	ABU58660
23	1924.5	68.8	527	24	ABU56196
24	1924.5	68.8	527	24	ABU57191
25	1924.5	68.8	527	24	ABU10770
26	1919.5	68.6	527	23	ABP71237
27	1913.5	68.4	527	24	ABU54566
28	1912	68.3	507	23	AAE16940
29	1874.5	67.0	454	23	ABG27193
30	1795.5	64.2	548	22	ABG27193
31	1699	60.7	533	23	ABG8414
32	1137.5	40.7	533	23	AAU10940
33	1137.5	40.7	533	23	AAU10941
34	1137.5	40.7	533	24	ABG99721
35	1120.5	40.0	530	23	ABG08490
36	1120.5	40.0	530	23	ABG08491
37	1120.5	40.0	530	23	ABG04429
38	1120.5	40.0	530	23	ABG04430
39	1120.5	40.0	533	13	AA26153
40	1114	39.8	544	12	AA26154
41	1091	39.0	544	22	ABG27065
42	1059	37.8	801	22	ABG27195
43	803	28.7	541	23	ABG57264
44	765.5	27.4	245	21	AAV57100
45	695	24.8	516	22	ABG62563

ALIGNMENTS

RESULT 1	
AA28677	
ID	AA28677 standard; Protein; 529 AA.
AC	AA28677;
DT	13-FEB-2001 (first entry)
XX	
DE	Human carbohydrate-modifying enzyme Incyte ID No: 2912330CD1.
XX	
KW	Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW	immunosuppressive; anti-HIV; antiinflammatory; antianemic;
KW	antiatherosclerotic; antihypertensive; antidiabetic;
KW	nephrotropic; antitumor; thymic; neuroprotective; osteoprotic;
KW	antiarthritic; antipruritic; uterine; ophthalmologic;
KW	dermatologic; antitumor; cytostatic; vitreous; antibacterial;
KW	fungicide; protozoicide; tranquilizer; vulnery; diabetes;
KW	autoimmune disorder; inflammatory disorder; infection.
XX	
OS	Homo sapiens.
XX	
PN	W0200063351-A2.
XX	
PD	26-OCT-2000.
XX	
PF	20-APR-2000; 2000MO-US10882.
XX	
PR	21-APR-1999; 99US-0130383.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Lai P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;

XX WPI: 2000-672729/65.
 DR N-PSDB; AAC65396.
 XX Novel carbohydrate modifying enzyme polypeptides and polynucleotides
 PT for diagnosis, treatment, and prevention of carbohydrate metabolism
 PT disorders, autoimmune/inflammatory disorders, and cancer
 XX
 PS Claim 1; Page 71-72; 75pp; English.
 XX
 CC The present sequence is a human carbohydrate-modifying enzyme
 CC (CME). CME polynucleotides and polypeptides are useful for treating and
 CC diagnosing diseases associated with CME such as diabetes,
 CC autoimmune/inflammatory disorders such as AIDS, Addison's disease,
 CC adult respiratory distress syndrome, allergies, anaemia, asthma,
 CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,
 CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
 CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
 CC osteoarthritis, osteoporosis, pancreatitis, polyomyelitis, psoriasis,
 CC Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic
 CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
 CC complications of cancer, haemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
 CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
 CC useful for drug screening.
 CC
 XX Sequence 529 AA;
 SQ

Query Match 100.0%; Score 2798; DB 21; Length 529;
 Best Local Similarity 100.0%; Pred. No. 5e-265;
 Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 DB 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 QY 61 SISDPNPSPTLKEFVYVSVSLTKTEFEDIKQLVKRAELPKDFTWSYFSQVOEIMTFN 120
 DB 61 SISDPNPSPTLKEFVYVSVSLTKTEFEDIKQLVKRAELPKDFTWSYFSQVOEIMTFN 120
 QY 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 QY 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPYVVMVSELSDDQMTFIERVKMIYVLYFEFWFOIFDMKKWDQYSEVL 240
 DB 181 EKHSGGLFPSPYVVMVSELSDDQMTFIERVKMIYVLYFEFWFOIFDMKKWDQYSEVL 240
 QY 241 GRPPTLSEFMKADILIRNWDFOFPHPLPNVFEVGLHCKPKKPLPKMEEEVVOSSG 300
 DB 241 GRPPTLSEFMKADILIRNWDFOFPHPLPNVFEVGLHCKPKKPLPKMEEEVVOSSG 300
 QY 301 ENGUVVFSIGSVWNTSEBRANVIASALAKIPQKYLWPFQGNKPDITGLNTRLYKMIQN 360
 DB 301 ENGUVVFSIGSVWNTSEBRANVIASALAKIPQKYLWPFQGNKPDITGLNTRLYKMIQN 360
 QY 361 DILGHPKTKAFTTHGANGIYEAIIHGVPMGVPIFGDGLDNIAMKAKGAVALNFKTM 420
 DB 361 DILGHPKTKAFTTHGANGIYEAIIHGVPMGVPIFGDGLDNIAMKAKGAVALNFKTM 420
 QY 421 TSEDLLRLARTVITDSSYKENAMRLSRIHDDQVPLDPAVWIFEFWRHKAHLSRA 480
 DB 421 TSEDLLRLARTVITDSSYKENAMRLSRIHDDQVPLDPAVWIFEFWRHKAHLSRA 480
 QY 481 HDLTFQHSIDVIGELTCAVATAIFLTCKFLFSQCKNKTKRIEKRE 529
 DB 481 HDLTFQHSIDVIGELTCAVATAIFLTCKFLFSQCKNKTKRIEKRE 529

RESULT 2
 AAY78933
 standard; Protein; 528 AA.

XX AAY78933;
 AC 05-JUN-2000 (first entry)
 DT Human UDP-glucuronosyltransferase 2B4 amino acid sequence.
 DE
 XX Human UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
 XX drug interaction; detect; human; single nucleotide polymorphism.
 XX
 OS Homo sapiens.
 OS
 XX MO200006776-A1.
 XX
 PD 10-FEB-2000.
 XX
 XX 22-JUL-1999; 99W0-US16675.
 XX
 PR 28-JUL-1998; 98US-0094391.
 XX
 PA (AXYS-) AXYS PHARM INC.
 XX
 PL Galvin M, Miller A, Penny L, Riedy M;
 XX
 DR WPI: 2000-195321/17.
 DR N-PSDB; AA295199.
 XX
 PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
 PT genotyping individuals to predict rate of metabolism of substrates and
 PT for identifying potential drug interactions
 PS Disclosure; Page 36-37; 72pp; English.
 XX

CC This sequence represents the human UDP-glucuronosyltransferase 2B4
 CC (UGT2B4) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
 CC family of enzymes that catalyze the glucuronic acid conjugation of a
 CC wide range of endogenous and exogenous substrates. The UGT2B gene
 CC subfamily encode steroid metabolizing isoforms in the liver. Alteration
 CC of the expression or function of UGTs may effect drug metabolism. The
 CC invention relates to non-chromosomal nucleic acid molecules, which
 CC comprise human UGT2B sequence polymorphisms (see AA295051-295110). Probes
 CC which detect the UGT2B locus polymorphisms can be used to detect altered
 CC UGT2B metabolism of a substrate in an individual. The nucleic acid
 CC molecules comprising a human UGT2B sequence polymorphism can be used in
 CC screening assays for genotyping individuals, also to predict their rate
 CC of metabolism of UGT2B substrate, potential drug-drug interactions and
 CC adverse side effects. The polymorphisms can be used as single nucleotide
 CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
 CC variation in activity or expression of UGT2B protein. The polymorphism
 CC containing nucleic acid molecules may also be used for generating
 CC genetically modified non-human animals and for obtaining site specific
 CC gene modification in cell lines.
 XX

Query Match 93.0%; Score 2602.5; DB 21; Length 528;
 Best Local Similarity 93.0%; Pred. No. 7.4e-246;
 Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;

QY 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 DB 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 QY 61 SISDPNPSPTLKEFVYVSVSLTKTEFEDIKQLVKRAELPKDFTWSYFSQVOEIMTFN 120
 DB 61 SISDPNPSPTLKEFVYVSVSLTKTEFEDIKQLVKRAELPKDFTWSYFSQVOEIMTFN 120
 QY 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYSNKKLKKLOESRPDVLA VFPFGEIALLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPYVVMVSELSDDQMTFIERVKMIYVLYFEFWFOIFDMKKWDQYSEVL 240
 DB 181 EKHSGGLFPSPYVVMVSELSDDQMTFIERVKMIYVLYFEFWFOIFDMKKWDQYSEVL 240

not applicable

Db 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMIYLYFEFFWQIFDMKKMDQFYSEVL 240
 Qy 241 GRPTLSETMAKADIWILIRNYWDPQPHLLPNVEFGGLHCKPAKPLPKMEEFVSSG 300
 Db 241 GRPTLSETMAKADIWILIRNYWDPQPHLLPNVEFGGLHCKPAKPLPKMEEFVSSG 300
 Qy 301 ENGVSFSLGSMVNTSEERANVIASALAKI POKVLMRFQDNKPDGLNTRLYKWI POK 360
 Db 301 ENGVSFSLGSMVNTSEERANVIASALAKI POKVLMRFQDNKPDGLNTRLYKWI POK 360
 Qy 361 DLGHPTKATITGGANNGIYEALYHGVPMVGVPIFGQDLNIAHMKAKGAAVEINFKTM 420
 Db 361 DLGHPTKATITGGANNGIYEALYHGVPMVGVPIFGQDLNIAHMKAKGAAVEINFKTM 420
 Qy 421 TSEDLLRALRTVITDSSYKENAMRLSRHHDPYKPLDRAVFWI EFWMRHKGAKHLRSAA 480
 Db 421 TSEDLLRALRTVITDSSYKENAMRLSRHHDPYKPLDRAVFWI EFWMRHKGAKHLRSAA 480
 Qy 481 HDLTFQHYSDIVIGFLTLCVATAIFLTKCFLSCQKFNKTRKIEKRE 529
 Db 481 HDLTFQHYSDIVIGFLTLCVATAIFLTKCFLSCQKFNKTRKIEKRE 529
 Db 481 HDLTFQHYSLDVTGFLACVATVIFITTC-LFCVWKFFVTRGKKGRD 528
 RESULT 3
 ID ABO5525 standard; Protein; 540 AA.
 XX ABO5525:
 AC ABO5525:
 XX 13-FEB-2002 (first entry)
 DT 13-FEB-2002 (first entry)
 XX Novel human diagnostic protein #5516.
 DE
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI, 2001-639362/73.
 DR N-PSDB; AAS69712.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT
 PS Claim 20; SEQ ID No 35884; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABO50010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 540 AA;
 Query Match 92.4%; Score 2584.5; DB 22; Length 540;
 Best Local Similarity 92.4%; Pred. No. 4,4e-244;
 Matches 489; Conservative 14; Mismatches 25; Indels 1; Gaps 1;
 Qy 1 MSKMTSALLLIOLSCYFSSGCGKVLWPTFESHMNNITIDELVQGHVTVLASSA 60
 Db 13 MSKMTSALLLIOLSCYFSSGCGKVLWPTFESHMNNITIDELVQGHVTVLASSA 72
 Qy 61 SISPDPSPTLKPEVYPVSLTKTEFEDITKOLYKRAELPKDTFWSYFOVOEIMWTFN 120
 Db 73 SISPDPSPTLKPEVYPVSLTKTEFEDITKOLYKRAELPKDTFWSYFOVOEIMWTFN 132
 Qy 121 DILRPFCKDIVSNKKLMKQLQESRFVVLADAVFPFEGELAEELKIPFVYSLRSPGYAI 180
 Db 133 DILRPFCKDIVSNKKLMKQLQESRFVVLADAVFPFEGELAEELKIPFVYSLRSPGYAI 192
 Qy 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMIYLYFEFFWQIFDMKKMDQFYSEVL 240
 Db 193 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMIYLYFEFFWQIFDMKKMDQFYSEVL 252
 Qy 241 GRPTLSETMAKADIWILIRNYWDPQPHLLPNVEFGGLHCKPAKPLPKMEEFVSSG 300
 Db 253 GRPTLSETMAKADIWILIRNYWDPQPHLLPNVEFGGLHCKPAKPLPKMEEFVSSG 312
 Qy 301 ENGVSFSLGSMVNTSEERANVIASALAKI POKVLMRFQDNKPDGLNTRLYKWI POK 360
 Db 313 ENGVSFSLGSMVNTSEERANVIASALAKI POKVLMRFQDNKPDGLNTRLYKWI POK 372
 Qy 361 DLGHPTKATITGGANNGIYEALYHGVPMVGVPIFGQDLNIAHMKAKGAAVEINFKTM 420
 Db 373 DLGHPTKATITGGANNGIYEALYHGVPMVGVPIFGQDLNIAHMKAKGAAVEINFKTM 432
 Qy 421 TSEDLLRALRTVITDSSYKENAMRLSRHHDPYKPLDRAVFWI EFWMRHKGAKHLRSAA 480
 Db 433 TSEDLLRALRTVITDSSYKENAMRLSRHHDPYKPLDRAVFWI EFWMRHKGAKHLRSAA 492
 Qy 481 HDLTFQHYSDIVIGFLTLCVATAIFLTKCFLSCQKFNKTRKIEKRE 529
 Db 493 HDLTFQHYSLDVTGFLACVATVIFITTC-LFCVWKFFVTRGKKGRD 540
 RESULT 4
 ID AAO22794 standard; Protein; 529 AA.
 XX AAO22794:
 AC AAO22794:
 XX 21-NOV-2002 (first entry)
 DT 21-NOV-2002 (first entry)
 XX
 XX Protein of drug metabolizing enzyme - 7493833CD1.
 DE
 XX Anti-HIV, antiarteriosclerotic, dermatological; cytostatic; thyromimetic;
 KM osteopapathic; thrombolytic; ophthalmological; antiinflammatory; hepatotropic;
 KM antidiarrhoeic; antiinflammatory; virocidic; immunogenic; autoimmune;
 KM drug metabolizing enzyme; DME; inflammatory; AIDS; atherosclerosis;
 KM contact dermatitis; cell-proliferative; cancer; cirrhosis; dwarfism;
 KM developmental; hypothyroidism; endocrine; osteoporosis; thrombosis;
 KM diabetes; glaucoma; keratitis; metabolic; hyperlipidaemia; diarrhoea;
 KM cystic fibrosis; gastrointestinal; gastroenteritis; liver; hepatitis;
 KM Reye's syndrome; exogenous compound; gene therapy; enzyme; human.

OS Homo sapiens.
 XX
 XX MO200266654-A2.
 XX
 XX 29-AUG-2002.
 PD
 PF 14-FEB-2002; 2002WO-US04918.
 XX
 XX 16-FEB-2001; 2001US-269643P.
 PR 23-FEB-2001; 2001US-271332P.
 PR 16-MAR-2001; 2001US-276767P.
 PR 06-APR-2001; 2001US-282077P.
 PR 19-APR-2001; 2001US-285447P.
 PR 27-APR-2001; 2001US-287060P.
 PR 03-MAY-2001; 2001US-288543P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Astronomoff A, Au-young J, Baughn MR, Ding L, Duggan BW;
 PI Forsythe J, Gietzen KJ, Griffin JH, Lee BA, Lu Y, Richardson TW;
 PI Ring HZ, Sanjanna MM, Swarnakar A, Walla NK, Warren BA, Xu Y;
 PI Yue H, Zabarjadian Y;
 XX
 DR MPI; 2002-674949/72.
 DR N-PSDB; AAL41490.
 XX
 XX New drug metabolizing enzymes (DME) useful for diagnosing, treating and
 PT preventing diseases or conditions associated with aberrant DME
 PT expression, e.g. cancer, AIDS, atherosclerosis, diabetes, glaucoma,
 PT hepatitis, osteoporosis -
 XX
 PS Claim 1; Page 150-151; 166pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising 12 fully
 CC defined sequences of 81-615 amino acids given in the specification; a
 CC naturally occurring amino acid sequence at least 90% identical to, having
 CC 81-599 amino acids, at least 93% identical to a sequence of 529 amino
 CC acids, or at least 97% identical to a sequence of 615 amino acids, all
 CC given in the specification; or a biologically active or immunogenic
 CC fragment of the polypeptide. The polypeptides and polynucleotides are
 CC useful in diagnosing, treating and preventing diseases or conditions
 CC associated with the decreased expression or overexpression of a drug
 CC metabolizing enzyme (DME), such as autoimmune/inflammatory (e.g. AIDS,
 CC atherosclerosis, contract dermatitis) cell-proliferative (e.g. cancer,
 CC cirrhosis), developmental (e.g. dwarfism, hypothyroidism), endocrine
 CC (e.g. osteoporosis, chromoblastoma, diabetes), eye (e.g. glaucoma,
 CC keratitis), metabolic (e.g. hyperlipidaemia, cystic fibrosis),
 CC gastrointestinal (e.g. gastroenteritis, diarrhoea), or liver (e.g.
 CC hepatitis, Reye's syndrome) disorders. These are also useful in assessing
 CC the effects of exogenous compounds on the expression of nucleic acid and
 CC amino acid sequences of DME. The DME or its fragments are useful in
 CC screening compounds for effectiveness as agonist or antagonist of the
 CC polypeptides, or in altering the expression of the target polynucleotide
 CC and compounds that specifically bind to or modulate the activity of the
 CC polypeptide. The microarray is useful in monitoring or measuring protein-
 CC protein interactions, drug-target interactions, and gene expression
 CC profiles. The polynucleotides of the invention can be used in gene
 CC therapy. This sequence represents a drug metabolizing enzyme protein of
 CC the invention.
 CC
 XX
 XX Sequence 529 AA;
 SO
 Query Match 89.5%; Score 2503; DB 23; Length 529;
 Best Local Similarity 87.7%; Pred. No. 4.2e-236;
 Matches 464; Conservative 29; Mismatches 36; Indels 0; Gaps 0;

QY 121 DILKPKCDIVSNKMLKKLQESRFDVVLADAVFPFGLLAEKLPVYSLRSPGYAI 180
 |||
 DB 121 DILKPKCDIVSNKMLKKLQESRFDVLPADALFPCELLAELEFNIPVYSLSPGYTF 180
 |||
 QY 181 EKHSGLLPFPSPYVPMVMSLSDQMTFIEKVKNNIYLYEFMFQIFDMKKMDQFYSEVL 240
 |||
 DB 181 EKHSGLFPFPSPYVPMVMSLSDQMTFIEKVKNNIYLYEFMFQIFDMKKMDQFYSEVL 240
 |||
 QY 241 GRPPTLSEMAKDIWIRBYWDPQPPHPLPVPVEFGGHCAPKPLPKMESEFVSSG 300
 |||
 DB 241 GRPPTLSEMAKDIWIRBYWDPQPPHPLPVPVEFGGHCAPKPLPKMESEFVSSG 300
 |||
 QY 301 ENGIVVPSLGSMTSNTSEERANVYASALAKIPKVLRFPGNKPDTLGLTRLYKWIPON 360
 |||
 DB 301 ENGIVVPSLGSMTSNTSEERANVYASALAKIPKVLRFPGNKPDTLGLTRLYKWIPON 360
 |||
 QY 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKAAVEINFTKM 420
 |||
 DB 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKAAVEINFTKM 420
 |||
 QY 421 TSEDLALRLTVITDSSYKENAMRLSRIHHDQPKPLDPAVFWTEFVRRHKGAKHLSSAA 480
 |||
 DB 421 TSEDLALRLTVITDSSYKENAMRLSRIHHDQPKPLDPAVFWTEFVRRHKGAKHLSSAA 480
 |||
 QY 481 HDLTFQHSIDVIGFLTCVATLIFLFTKCFPLFSCQFNKTRKIEKRE 529
 |||
 DB 481 HDLTFQHSIDVIGFLTCVATLIFLFTKCFPLFSCQFNKTRKIEKRE 529
 |||
 RESULT 5
 AAE28617
 ID AAE28617 standard; Protein; 529 AA.
 XX
 AC AAE28617;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 XX Human UGT2B7 protein.
 XX
 XX Human; UDP-glucuronosyl transferase; UGT; UGT2B7; toxicity; cancer;
 KW therapy; epirubicin; cytosol; enzyme.
 XX
 OS Homo sapiens.
 XX
 XX MO200259375-A2.
 XX
 XX 01-AUG-2002.
 XX
 PD 25-JAN-2002; 2002WO-US02083.
 XX
 PF 26-JAN-2001; 2001US-264534P.
 XX
 PR (UYCH-) UNIV CHICAGO.
 PA
 PI Ratain MJ, Innocenti F, Das S, Iyer L, Sawyer M;
 XX
 XX MPI; 2002-691534/74.
 DR N-PSDB; AAD45991.
 XX
 XX Determining the dose of a UGT2B7-glucuronidated drug for treating
 PT cancer, comprises determining the level of UGT2B7 activity or
 PT expression in a patient -
 XX
 PS Disclosure; Page 144-145; 160pp; English.
 XX
 CC The invention relates to a UDP-glucuronosyl transferase (UGT) enzyme,
 CC UGT2B7. The invention also relates to compositions and methods for
 CC optimising UGT2B7 substrate dosings and for predicting UGT2B7 substrate
 CC toxicity. The method is useful in determining the dose of a UGT2B7-
 CC glucuronidated drug that may be used in treating cancer patients. It
 CC is also useful in determining persons at risk for epirubicin toxicity,
 CC in reducing or eliminating side effects associated with epirubicin

CC treatment, and in ways of increasing the efficacy of dosage regimens.
CC The present sequence is human UG72B7 protein.

XX Sequence 529 AA;

Query Match 83.3%; Score 2330; DB 23; Length 529;
Best Local Similarity 81.1%; Pred. No. 3.9e-219;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```
QY 1 MSKMTSALLILQLSCFSSGCGKVLVMPTEFSHMNNTITLDELVORGEHTVLASSA 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSVMTSVILILQLSCFSSGCGKVLVMAEYSHMNMKITLDELVORGEHTVLASSA 60
QY 61 SIFDPPSPSTLKEEYVPVSLTKTEFEDIKOLVKMAELPKOTFMSYFSQVOEIMWTFN 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SIFDPPNNSALKIEIYPTSLTKLENFIMQIKRNSDLPKOTFMYLFSQVOEIMNIFG 120
QY 121 DILRKFCDDIVSNKKMKKLQESRPDVVLADAVPFGELLAELIKIPVYSLRFSGYAI 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 DILRKFCDDIVSNKKMKKLQESRPDVVLADAVPFGELLAELIKIPVYSLRFSGYAI 180
QY 181 EKHSGGILPPSYVPVVMSELSDQMTFIERVKMNIYVLFYEFWFOITDMKKMDQFYSEVL 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EKHSGGILPPSYVPVVMSELSDQMTFIERVKMNIYVLFYEFWFOITDMKKMDQFYSEVL 240
QY 241 GRPTLSETMAKADIMILIRNYMDFOPHPLLPNVEFVGGLHCKPAKPLPKEMEFPVQSSG 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 GRPTLSETMAKADIMILIRNSMNFQPHPLLPNDVFGGLHCKPAKPLPKEMEFPVQSSG 300
QY 301 ENGVVVFSLGSMVSNTESEERANVIASALAKIPOKVLMPFGNKPDTLGLNTRLYKWIPON 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 ENGVVVFSLGSMVSNTESEERANVIASALAKIPOKVLMPFGNKPDTLGLNTRLYKWIPON 360
QY 361 DLIGHPKTKAFITGGGNGIYEALHYGVPMVGPIIFGDDQDNIAMKAKAAVEINIKTM 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 DLIGHPKTKAFITGGGNGIYEALHYGVPMVGPIIFGDDQDNIAMKAKAAVEINIKTM 420
QY 421 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFWIEFMRKHGAHLRSAA 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFWIEFMRKHGAHLRSAA 480
QY 481 HDLTFWFOHYSIDVIGFLITCVATAIPLFTKCFLFSCKQFNKTRKIEKRE 529
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 HDLTFWFOHYSIDVIGFLITCVATAIPLFTKCFLFSCKQFNKTRKIEKRE 529
```

RESULT 6
ABG05523
ID ABG05523 standard; Protein; 533 AA.

```
XX ABG05523;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #5514.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX
```

XX WI: 2001-639362/73.
DR N-PSDB; AAS69710.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 35882; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 533 AA;

Query Match 83.1%; Score 2325; DB 22; Length 533;
Best Local Similarity 80.9%; Pred. No. 1.2e-218;
Matches 428; Conservative 43; Mismatches 58; Indels 0; Gaps 0;

```
QY 1 MSKMTSALLILQLSCFSSGCGKVLVMPTEFSHMNNTITLDELVORGEHTVLASSA 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 MSVMTSVILILQLSCFSSGCGKVLVMAEYSHMNMKITLDELVORGEHTVLASSA 64
QY 61 SIFDPPSPSTLKEEYVPVSLTKTEFEDIKOLVKMAELPKOTFMSYFSQVOEIMWTFN 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 SIFDPPNNSALKIEIYPTSLTKLENFIMQIKRNSDLPKOTFMYLFSQVOEIMNIFG 124
QY 121 DILRKFCDDIVSNKKMKKLQESRPDVVLADAVPFGELLAELIKIPVYSLRFSGYAI 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 DILRKFCDDIVSNKKMKKLQESRPDVVLADAVPFGELLAELIKIPVYSLRFSGYAI 184
QY 181 EKHSGGILPPSYVPVVMSELSDQMTFIERVKMNIYVLFYEFWFOITDMKKMDQFYSEVL 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 EKHSGGILPPSYVPVVMSELSDQMTFIERVKMNIYVLFYEFWFOITDMKKMDQFYSEVL 244
QY 241 GRPTLSETMAKADIMILIRNYMDFOPHPLLPNVEFVGGLHCKPAKPLPKEMEFPVQSSG 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 GRPTLSETMAKADIMILIRNSMNFQPHPLLPNDVFGGLHCKPAKPLPKEMEFPVQSSG 304
QY 301 ENGVVVFSLGSMVSNTESEERANVIASALAKIPOKVLMPFGNKPDTLGLNTRLYKWIPON 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 ENGVVVFSLGSMVSNTESEERANVIASALAKIPOKVLMPFGNKPDTLGLNTRLYKWIPON 364
QY 361 DLIGHPKTKAFITGGGNGIYEALHYGVPMVGPIIFGDDQDNIAMKAKAAVEINIKTM 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 365 DLIGHPKTKAFITGGGNGIYEALHYGVPMVGPIIFGDDQDNIAMKAKAAVEINIKTM 424
QY 421 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFWIEFMRKHGAHLRSAA 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 425 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFWIEFMRKHGAHLRSAA 484
QY 481 HDLTFWFOHYSIDVIGFLITCVATAIPLFTKCFLFSCKQFNKTRKIEKRE 529
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db	485	HDLTMFQIHSDDVIGFLVCATVTFIVTKCCLPFCWKFARAKKXND	513
RESULT 7			
ID	AAY78934		
AC	AAY78934	standard; Protein; 524 AA.	
XX			
DT	05-JUN-2000	(first entry)	
XX			
DE		Human UDP-glucuronosyltransferase 2B7 amino acid sequence.	
XX			
KM		UDP-glucuronosyltransferase 2B8; UGT2B4; polymorphism; metabolism; SNPs;	
XX		drug interaction; detect; human; single nucleotide polymorphism.	
OS			
XX			
XX		Homo sapiens.	
PN			
XX		WO200006776-A1.	
PD			
XX		10-FEB-2000.	
PF			
XX		22-JUL-1999; 99WO-US16675.	
PR			
XX		28-JUL-1998; 98US-0094391.	
PA			
PI		(AXYS-) AXYS PHARM INC.	
DR		Galvin M, Miller A, Penny L, Riedy M;	
XX		MPi; 2000-195321/17.	
DR		N-PsDB; AA95200.	
PT			
PS			
XX			
XX		Noel human UDP-glucuronosyltransferase sequence, polymorphisms for	
XX		genotyping individuals to predict rate of metabolism of substrates and	
XX		for identifying potential drug interactions -	
XX			
XX		Disclosure; Page 44-45; 72pp; English.	
CC			
CC		This sequence represents the human UDP-glucuronosyltransferase 2B7	
CC		(UGT2B7) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a	
CC		family of enzymes that catalyse the glucuronic acid conjugation of a	
CC		wide range of endogenous and exogenous substrates. The UGT2B gene	
CC		subfamily encode steroid metabolizing isoforms in the liver. Alteration	
CC		of the expression or function of UGTs may effect drug metabolism. The	
CC		invention relates to non-chromosomal nucleic acid molecules, which	
CC		comprise human UGT2B sequence polymorphisms (see AA295051-Z95110). Probes	
CC		which detect the UGT2B locus polymorphisms can be used to detect altered	
CC		UGT2B metabolism of a substrate in an individual. The nucleic acid	
CC		molecules comprising a human UGT2B sequence polymorphism can be used in	
CC		screening assays for genotyping individuals, also to predict their rate	
CC		of metabolism of UGT2B substrate, potential drug-drug interactions and	
CC		adverse side effects. The polymorphisms can be used as single nucleotide	
CC		polymorphisms (SNPs) for detecting genetic linkage related to phenotypic	
CC		variation in activity or expression of UGT2B protein. The polymorphism	
CC		containing nucleic acid molecules may also be used for generating	
CC		genetically modified non-human animals and for obtaining site specific	
CC		gene modification in cell lines.	
SQ			
Sequence	524 AA;		
Query Match	83.0%; Score 2322; DB 21; Length 524;		
Best Local Similarity	81.5%; Pred. Mismatches 2, 3e-218;		
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0			
OY	1 MSWKTSALLLIQLSCYFSSGSGCGKYLWPEFSHMNTKITIDELVORCHEVTYLAASA 60		
DB	1 MSVKMTSVILLIQLSCFSSGNCGKYLVAAEYSHMNITITIDEILIORHEVTYLAASA 60		
OY	61 SISFDNSNSTLKFEYVPVSUTTEBEDIIKOLYKMAEIPKOTFMWSYFSQVEINMTEN 120		
DB	61 SLTPFDNNSSALKETIEYPTSLTTELLENFTMOQIKMSDLPKOTFMWYFSQVEINSIFG 120		

Qy	12	DIIRKXCIOIVSNKKUMKMGLOSREDVYLAADVPCGGLLAELIKIIPVYSLSEFS	PGYAI	180
Db	121	DIIRKCKXOVSNKFKMKVQDSRDVIFADAIIFPCSELALFNIFVYSLSEFS	PGYTF	180
Qy	181	EKISGGILFPSPSVPVWVSELSDOQTFIERVKMIIYVLYEFWFQIFDMKKMDQ	YSEVL	240
Db	181	EKISGGFIIPPSVVPVWVSELTDOQTFEMERYKMIIYVLYFDFWPEIFDMKKMDQ	YSEVL	240
Qy	241	GREPTLSEFMKADIIWIRNYWDPQFPHPDLLPNVEFVGGHCKRAKYLPRKME	EPVQSSG	300
Db	241	GREPTLSEFMKADYVLIIRNSWNPFPYPPLLPNDVPVGGHCKRAKPLPRME	EDPVQSSG	300
Qy	301	ENGVVVFISGSWVNSTSEERAVIVASALAKIPOKYLMPFDNKEDTGLTRILYK	WPIPON	360
Db	301	ENGVVVFISGSWVNSTSEERAVIVASALAQIPOKYLMPFDNKEDTGLTRILYK	WPIPON	360
Qy	361	DLGHPKRTKAFITHGGMNGIYEAIHGVPMVGPVIFGQDLDNIAMKAKAGA	VEINFTKM	420
Db	361	DLGHPKRTKAFITHGGMNGIYEAIHGIIPMGIFLFAQDPNIIAMKAKAGA	VAVRDENTM	420
Qy	421	TSDELLRALRTVITDSSYKENAMRLSRIHHDDPVKPELDRAVFWIEFVNRH	KGAHLRSLA	480
Db	421	TSDELLNALAKRVINDPSSYKENAMKLSRIQHDQPVKPELDRAVFWIEFVNRH	KGAHLRSLA	480
Qy	481	HDITWFOHYSIDVIGLITCVATALFLTKCLJPSCKQENKTRK	524	
Db	481	HDITWFOHSLDVIGLITCVATVFIYTKCCLEFEMWFAKAK	524	

RESULT 8
 AAE02188
 ID AAE02188 standard; Protein; 529 AA.
 XX
 AC AAE02188;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human breast cancer specific gene-2 (BCSG-2) protein.
 XX
 KM Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
 XX breast cancer; therapeutic; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200137779-A2.
 XX
 PD 31-MAY-2001.
 XX
 PF 22-NOV-2000; 2000MO-US32056.
 XX
 PR 23-NOV-1999; 99US-0166973.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Salceda S, Cafferty R, Recipon H, Sun Y;
 XX
 DR WPI: 2001-367602/38.
 XX
 DR N-PSDB; AAD06860.
 XX
 PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
 XX imaging, preventing and treating cancers, particularly breast cancer
 XX
 PS Claim 2; Page 55-57; 66pp; English.
 CC
 CC The invention relates human breast cancer specific genes (BCSG's) and
 CC their corresponding proteins. BCSG is useful for diagnosing, staging,
 CC monitoring, imaging, preventing and treating breast cancers. BCSG is also
 CC useful for inducing an immune response against a target cell expressing
 CC BCSG. The invention also provide methods for detecting genetic lesions or
 CC mutations in BCSG, thereby determining if a human with the genetic lesion
 CC is at risk for breast cancer or has breast cancer. BCSG antibodies
 CC labelled with paramagnetic ions or radioisotopes is useful for imaging
 CC breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is

CC useful for treating breast cancer. BCSG is useful in the rational design
 CC of new therapeutics for imaging and treating cancers. BCSG is also used
 CC in gene therapy. The present sequence is human breast cancer specific
 CC gene-2 (BCSG-2) protein.

XX Sequence 529 AA;

Query Match 81.9%; Score 2292; DB 22; Length 529;
 Best Local Similarity 80.2%; Pred. No. 2e-215;
 Matches 424; Conservative 43; Mismatches 62; Indels 0; Gaps 0;

```

QY 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTLASSA 60
DB 1 MTLKMTSVLLIHLISCFYSSGCGKVLVMAEYSHMMNMKITLKLVLQRGHEVTLASSA 60
QY 61 SISFDNPSPTLKEVYVPSLTKTEFEDIIKOLVYKMAELIKDQFWSYFSQVQEIIMTFN 120
DB 61 SILFDPNASTLKEVYVPSLTKTEFENIMQOVYRMSDIKRSDFWLSFQOEILWELV 120
QY 121 DILKPKCDIYVSNKKLMKKLOESRFDVLADAVFPFGSLAELLKIPVYSLRSPGYAI 180
DB 121 DIFNPKCDIYVSNKKLMKKLOESRFDIYFADAVFPFGSLAELLKIPVYSLRFTPGYTI 180
QY 181 EKHSGLIFPPSYVYVMSLSQMTFIERVKNMIVLYFEFPFQIFDMKKMDQFYSEVL 240
DB 181 ERHSGGLIFPPSYIPIVMSKLSQMTFIERVKNMIVLYFPFPMQMSDKMDQFYSEVL 240
QY 241 GRPTTSETMAKADIVLIRNWDQFPHPLPNVEFVGGLHCKRAKPLPKMEEFVQSSG 300
DB 241 GRPTTLEETMAKADIVLIRNWSQFPHPLPNVDFVGGFHCKRAKPLPKMEEFVQSSG 300
QY 301 ENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLMEFDGNKPTGLNTRLYKMIPON 360
DB 301 ENGIVVFSLSGMSVNTSEERANVIATALAKIPQKVLMEFDGNKPTGLNTRLYKMIPON 360
QY 361 DILGHPKTRAFITGGANGIYEAIYHGVPMVGPVIFGDQDNLIAHMKAGAAVEINFKTM 420
DB 361 DILGHPKTRAFITGGANGIYEAIYHGVPMVGPVIFGDQDNLIAHMKAGAAVLDFTMT 420
QY 421 TSEDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRSA 480
DB 421 SSTDLNALKRTVINDPXYKENIMKLSRIQHDQPVKPLDRAVFWIEFVWRHKGAKHLRSA 480
QY 481 HDLTFWFOHSIDVIGFLTCAATAPLFTKCFELSCOKENKTRIKRE 529
DB 481 HDLTFWFOHSIDVIGFLTCAATAPLFTKCFELSCOKENKTRIKRE 529

```

RESULT 9

ABG05524 ID ABG05524 standard; Protein; 532 AA.

AC ABG05524;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5515.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX DRmanac RT, Liu C, Tang YT;
 XX WI: 2001-639362/73.
 DR N-PSDB: AAS69711.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID No 35683; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes.
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 532 AA;

Query Match 79.8%; Score 2233; DB 22; Length 532;
 Best Local Similarity 79.4%; Pred. No. 1.3e-209;
 Matches 421; Conservative 43; Mismatches 64; Indels 2; Gaps 2;

```

QY 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTLASSA 60
DB 4 MALKMTT-VLLIQSFSSGCGKVLVMAEYSHMMNMKITLKLVLQRGHEVTLASSA 62
QY 61 SISFDNPSPTLKEVYVPSLTKTEFEDIIKOLVYKMAELIKDQFWSYFSQVQEIIMTFN 120
DB 63 SILFDPNDSSTLKEVYVPSLTKTEFENIMQOVYRMSDIKRSDFWLSFQOEILWELV 122
QY 121 DILKPKCDIYVSNKKLMKKLOESRFDVLADAVFPFGSLAELLKIPVYSLRSPGYAI 180
DB 123 DILKPKCDIYVSNKKLMKKLOESRFDIYFADAVFPFGSLAELLKIPVYSLRFTPGYTI 182
QY 181 EKHSGLIFPPSYVYVMSLSQMTFIERVKNMIVLYFEFPFQIFDMKKMDQFYSEVL 240
DB 183 ERHSGGLIFPPSYVYVMSLSQMTFIERVKNMIVLYFPFPMQMSDKMDQFYSEVL 242
QY 241 GRPTTSETMAKADIVLIRNWDQFPHPLPNVEFVGGLHCKRAKPLPKMEEFVQSSG 300
DB 243 GRPTTSETMAKADIVLIRNWSQFPHPLPNVDFVGGFHCKRAKPLPKMEEFVQSSG 302
QY 301 ENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLMEFDGNKPTGLNTRLYKMIPO 359
DB 303 ENGIVVFSLSGMSVNTSEERANVIATLAKIPQKVLMEFDGNKPTGLNTRLYKMIPO 362
QY 360 NDILGHPKTRAFITGGANGIYEAIYHGVPMVGPVIFGDQDNLIAHMKAGAAVEINFKT 419
DB 363 NDILGHPKTRAFITGGANGIYEAIYHGVPMVGPVIFGDQDNLIAHMKAGAAVLDFTMT 422
QY 420 MTSDDLRLALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRSA 479
DB 423 MSTDLNALKRTVINDPXYKENIMKLSRIQHDQPVKPLDRAVFWIEFVWRHKGAKHLRSA 482

```

Oy 480 AHDLTWFQHSIDVIGFLITCVATAIFLFTKCFLFSQCKFNKTRKIEKRE 529
 Db 483 AHNLTWQYHSIDVIGFLITCVATAIFLFTKCFLFSQCKFNKTRKIEKRE 532

RESULT 10

AAW47126 standard; Protein; 530 AA.

AAW47126;

26-MAY-1998 (first entry)

Uridine diphospho-glucuronosyltransferase 2B17 (UGT2B17) enzyme.

Uridine diphospho-glucuronosyltransferase 2B17; UGT2B17; catalyze;

androstereone; androstereone-glucuronic acid; androgen; enzyme.

Homo sapiens.

WO9744466-A1.

27-NOV-1997.

16-MAY-1997; 97WO-CA00328.

17-MAY-1996; 96US-0649319.

(ENDO-) ENDORECHERHE INC.

Beaulieu M, Belanger A, Hum DW, Levesque E;

WPI; 1998-018520/02.

N-PSDB; AAV15900.

DNA encoding uridine di:phospho:glucuronosyl:transferase 2B17 -

which catalyses conversion of androstereone to

androstereone-glucuronic acid

Claim 16; Pages 4-6; 53pp; English.

This is the enzyme uridine diphospho-glucuronosyltransferase 2B17

(UGT2B17). This novel enzyme catalyses the conversion of androstereone

to androstereone-glucuronic acid. The UGT2B17 can be used to detect

anti-UGT2B17 antibodies. The antibody can be used to detect a localised

concentration of UGT2B17 or an alteration in androgen activity. The

UGT2B17 can also be used to alter the concentration of an androgenic

compound in a tissue, specifically dihydrotestosterone. An isolated

nucleotide sequence comprising at least 30 consecutive nucleotides from

the coding region of the 2107 base pair sequence, or its complement can

be used to block the synthesis of UGT2B17, e.g. an expression disrupting

sense or antisense fragment, or as a probe for a UGT2B17 coding sequence.

Sequence 530 AA;

Query Match 77.5%; Score 2169.5; DB 19; Length 530;

Best Local Similarity 74.7%; Pred. No. 2.1e-203;

Matches 396; Conservative 60; Mismatches 73; Indels 1; Gaps 1;

1 MSMKTSALLIQLSCYFSSGSCGKVLVPTFESHMNIKITLDELVQGHSEVTLASSA 60
 1 MSKMSVFLMLQSLCYFSSGSCGKVLVPTFESHMNIKITLDELVQGHSEVTLASSA 60
 61 SISFSPNSPSTLKFVYVPSLTKTEPEDITIKOLVKRMA-ELPKDTFWSYFSCVOEIMTFE 119
 61 SILVAVASSSAIKLEIYVPTSLTKNLEDPFMKPFMRWTYSISKNTFWSYFSQLQELCWEY 120
 120 NDILAKFCDDIYSNKKMKKLOESRPDVVLAADVPFGELIAELKIPPVLSLRSPGA 179
 121 SDYNIKTCEDAVLNKKMKKLOESKRPDVVLAADVPFGELIAELNIPPLYSLRSPVGT 180
 180 IEKHSGGLTFPPSYVPVWSELSDQMTFERVKNMIVVYFFFWFOIFPMKKMDQFYSEV 239

Db 181 VEKNGGFLFPPSYVPVWSELSDQMTFERVKNMIVVYFFFWFOYDLKKMDQFYSEV 240
 Oy 240 LGPRTTSETMAKADILIRNYWDFOPPHLPVVEFVGJHCKPAKPLPEMEBFVQSS 299
 Db 241 LGPRTTSETMAKADILIRNYWDFOPPHLPVVEFVGJHCKPAKPLPEMEBFVQSS 300
 Oy 300 GENGWVFSLSGWSVNTSEERANVIALAKIPQKVLMPDGNKPDITGLNTRLYKWIPO 359
 Db 301 GENGWVFSLSGWSVNTSEERANVIALAKIPQKVLMPDGNKPDITGLNTRLYKWIPO 360
 Oy 360 NDILGHPKTKAFITHGNGNGIYENIYGVPMVGPITGDDLDNIAHKKAKGALEINFKT 419
 Db 361 NDILGHPKTKAFITHGNGNGIYENIYGVPMVGPITGDDLDNIAHKKAKGALEINFKT 420
 Oy 420 MTEEDLRALRTVITDSSYKENAMRLSRHHDDQVKKPLDRAVFIIEFMRHKGAKLRS 479
 Db 421 MSSRDLNALSKSVNDPIYENIMKLSRIHDDQVKKPLDRAVFIIEFMRHKGAKLRS 480
 Oy 480 AHDLTWFQHSIDVIGFLITCVATAIFLFTKCFLFSQCKFNKTRKIEKRE 529
 Db 481 AHNLTWQYHSIDVIGFLITCVATAIFLFTKCFLFSQCKFNKTRKIEKRE 530

RESULT 11

AA78935 standard; Protein; 530 AA.

AA78935;

05-JUN-2000 (first entry)

Human UDP-glucuronosyltransferase 2B15 amino acid sequence.

UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;

drug interaction; detect; human; single nucleotide polymorphism; SNPs.

Homo sapiens.

WO200006776-A1.

10-FEB-2000.

22-JUL-1999; 99WO-US16675.

28-JUL-1996; 98US-0094391.

(AAYS-) AAYS PHARM INC.

Galvin M, Miller A, Penny L, Riedy M;

WPI; 2000-195321/17.

N-PSDB; AA295206.

Novel human UDP-glucuronosyltransferase sequence, polymorphisms for

genotyping individuals to predict rate of metabolism of substrates and

for identifying potential drug interactions

Disclosure: Page 59-60; 72pp; English.

This sequence represents the human UDP-glucuronosyltransferase 2B15

(UGT2B15) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a

family of enzymes that catalyze the glucuronic acid conjugation of a

wide range of endogenous and exogenous substrates. The UGT2B gene

subfamily encode steroid metabolizing isoforms in the liver. Alteration

of the expression or function of UGTs may effect drug metabolism. The

invention relates to non-chromosomal nucleic acid molecules, which

comprise human UGT2B sequence polymorphisms (see AA295051-295110). Probes

which detect the UGT2B locus polymorphisms can be used to detect altered

UGT2B metabolism of a substrate in an individual. The nucleic acid

molecules comprising a human UGT2B sequence polymorphism can be used in

screening assays for genotyping individuals, also to predict their rate

of metabolism of UGT2B substrate, potential drug-drug interactions and

adverse side effects. The polymorphisms can be used as single nucleotide

CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
 CC variation in activity or expression of UGT2B protein. The polymorphism
 CC containing nucleic acid molecules may also be used for generating
 CC genetically modified non-human animals and for obtaining site specific
 CC gene modification in cell lines.

XX Sequence 530 AA;

Query Match 77.1%; Score 2158.5; DB 21; Length 530;
 Best Local Similarity 74.5%; Pred. No. 2.5e-202; Indels 1; Gaps 1;
 Matches 395; Conservative 60; Mismatches 74;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVMPTEFSHMNNTITLDELVORGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVMPTEFSHMNNTITLDELVORGHEVTVLASSA 60
 QY 61 SISDPNSPTLKEPVVSLTKTEFEDIIKQLYKRA-ELPKDTFMSYSQVQEIIMWTF 119
 DB 61 STLVAASKSAIKLEVPTSLTKNDLSDLSLKILDRMIVGVSKTFMSYSQVQELCWEY 120
 QY 120 NDILRKFKDIVSNKKLQESRFDVLADAVFPPEGLAEELKIPFVYSLRFSFGYA 179
 DB 121 YDYSNKLCKDAVLNKKLMLQESKFDVILADALNPGCELLAEELFNPLSLRFSVGYT 180
 QY 180 IEKHSGLLPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFVQIFDMKKMDQFYSEV 239
 DB 181 FEKKGGLFPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFVQIFDMKKMDQFYSEV 240
 QY 240 LGRPTTLESTMAKADILIRNYWDFOPPHLLPVEVEVGLHCKPAKLPKEMEFPVSS 299
 DB 241 LGRPTTLESTMAKADILIRNYWDFOPPHLLPVEVEVGLHCKPAKLPKEMEFPVSS 300
 QY 300 GENGIVVFSLSGMSVNTSEERANVYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 359
 DB 301 GENGIVVFSLSGMSVNTSEERANVYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 360
 QY 360 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAVEINFT 419
 DB 361 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAALSVDIRT 420
 QY 420 MTSDDLRLARTVITDSSYKENAMRLSRHHQDPVKPLDRAVPWIEFVMRHKGAKHLRSA 479
 DB 421 MSSRDLNLAKSVINDPVYKENVMKLSRIHHDQPMKPLDRAVPWIEFVMRHKGAKHLRSA 480
 QY 480 AHDLTWFOHYSIDVIGFLTGVATAIFLTKCFPLFSCQKFNKTRKIEKRE 529
 DB 481 AHNLTWIOHSLDVIAFLACVATVITITKFCFLCFRKLAKTKGKKRD 530

RESULT 12

ABJ19806
 ID ABJ19806 standard; Protein; 530 AA.

XX ABJ19806;

XX 10-APR-2003 (first entry)

DB Androgen-independent prostate cancer-related protein - SEQ ID No 12.

XX Androgen-independent cancer; androgen ablation therapy; prostate cancer;
 KW androgen-dependent prostate cancer; prostate cancer.

XX Unidentified.

XX WO200298358-A2.

XX 12-DEC-2002.

XX 04-JUN-2002; 2002WO-US17594.

XX 04-JUN-2001; 2001US-295917P.

PR 13-NOV-2001; 2001US-350666P.

PR 29-MAR-2002; 2002US-368689P.

PR 12-APR-2002; 2002US-372246P.
 PR 31-MAY-2002; 2002US-0160233.
 XX (BOSB-) EOS BIOTECHNOLOGY INC.
 PA Afar DEH, Agus D, Mack DH;
 XX WPI; 2003-148602/14.

PT Detecting an androgen-independent prostate cancer cell in a sample or
 PT diagnosing androgen-dependent prostate cancer, by determining the
 PT presence or absence of genes whose expressions are up- or
 PT down-regulated

PS Claim 1; Page 188; 21opp; English.

XX The invention comprises a method for detecting an androgen-independent
 CC cancer cell in a sample from a patient who has undergone androgen
 CC ablation therapy. The method involves determining the presence or absence
 CC of nucleic acids that are either up-regulated or down-regulated in
 CC prostate cancer. The method is useful for detecting an androgen-
 CC independent prostate cancer cell in a sample from a patient who has
 CC undergone androgen ablation therapy. The method is particularly useful
 CC for diagnosing androgen-dependent prostate cancer, prostate cancer
 CC undergoing androgen withdrawal, or androgen-independent prostate cancer.
 CC The present amino acid sequence represents a protein which is encoded by
 CC a gene that is either up-regulated or down-regulated in prostate cancer.

XX Sequence 530 AA;

Query Match 77.1%; Score 2158.5; DB 24; Length 530;
 Best Local Similarity 74.5%; Pred. No. 2.5e-202;
 Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVMPTEFSHMNNTITLDELVORGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVMPTEFSHMNNTITLDELVORGHEVTVLASSA 60
 QY 61 SISDPNSPTLKEPVVSLTKTEFEDIIKQLYKRA-ELPKDTFMSYSQVQEIIMWTF 119
 DB 61 STLVAASKSAIKLEVPTSLTKNDLSDLSLKILDRMIVGVSKTFMSYSQVQELCWEY 120
 QY 120 NDILRKFKDIVSNKKLQESRFDVLADAVFPPEGLAEELKIPFVYSLRFSFGYA 179
 DB 121 YDYSNKLCKDAVLNKKLMLQESKFDVILADALNPGCELLAEELFNPLSLRFSVGYT 180
 QY 180 IEKHSGLLPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFVQIFDMKKMDQFYSEV 239
 DB 181 FEKKGGLFPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFVQIFDMKKMDQFYSEV 240
 QY 240 LGRPTTLESTMAKADILIRNYWDFOPPHLLPVEVEVGLHCKPAKLPKEMEFPVSS 299
 DB 241 LGRPTTLESTMAKADILIRNYWDFOPPHLLPVEVEVGLHCKPAKLPKEMEFPVSS 300
 QY 300 GENGIVVFSLSGMSVNTSEERANVYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 359
 DB 301 GENGIVVFSLSGMSVNTSEERANVYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 360
 QY 360 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAVEINFT 419
 DB 361 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAALSVDIRT 420
 QY 420 MTSDDLRLARTVITDSSYKENAMRLSRHHQDPVKPLDRAVPWIEFVMRHKGAKHLRSA 479
 DB 421 MSSRDLNLAKSVINDPVYKENVMKLSRIHHDQPMKPLDRAVPWIEFVMRHKGAKHLRSA 480
 QY 480 AHDLTWFOHYSIDVIGFLTGVATAIFLTKCFPLFSCQKFNKTRKIEKRE 529
 DB 481 AHNLTWIOHSLDVIAFLACVATVITITKFCFLCFRKLAKTKGKKRD 530

RESULT 13

AAU29284

ID AAU29284 standard; Protein; 527 AA.
XX
AC AAU29284;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human PRO polypeptide sequence #261.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191114P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193033P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199507P.
PR 25-APR-2000; 2000US-199549P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209632P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-064484P.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR MPI; 2001-602746/68.
DR N-PSDB; AAS46185.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
XX presence of tumours, such as prostate and breast tumours, in mammals and
XX to screen for modulators of the compounds -
XX

PS Claim 11; Fig 522; 774bp; English.
XX
XX Sequences AAU29284-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX
SQ Sequence 527 AA;
Query Match 68.8%; Score 1924.5; DB 22; Length 527;
Best Local Similarity 69.9%; Pred. No. 2.2e-179;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
QY 2 SMKWTSAALLIQLSCYFSSGCGKLVWPTFESHMNIKITLDELVORGHEVTLASSAS 61
DB 3 SDRKALVFLQLQLFC-VGCGRCGKLVWPCDMSHMLAVKYLLEELIVRGHEVTLTHSKP 61
QY 62 ISFDNPSPTLKFEVYVSLTKTEFEDIIKQVYKMAELPRDPTWSYFSQOELMTWTFND 121
DB 62 SLIDYRKPSALKFEVYVHMPDRTEENEIFVDLALN--VLPELSTWQSVIKLNDFFVEIRG 119
QY 122 ILRFPCKDIVSNKRLMKLOESRFDVVLADAVPFGELAEILKIPFYVSRFSGYALE 181
DB 120 TLKWCSSFIYNQTLMKLOETNYDVMILDIVLPCGDLMAELLAVPFLVLRISVGGNME 179
QY 182 KHSGLLFPSPYVNVNSELSDQMTFERVNMIVLVYFEFWFOIPDMKKWDQFSEVTLG 241
DB 180 RSCGKLAPLSLYVVPMTGLDRMTFLERVNSMLSLVLFHWIMIDYDHFMEEFYSKALG 239
QY 242 RPTLSETMAKADILIRNWDPOFPHPLDNVEFVGGHCKPAKPLPKEMEETVQSSGE 301
DB 240 RPTLCETVQKAEIWLIRTYWDFEFPPOYPQNPFEVGGHCKPAKALPKEMENFVQSSGE 299
QY 302 NGVVVFSLSGWSVNTSEBRANVIVASALAKIPQKLVMPDGGKPTLGLNTRLYKWIPOND 361
DB 300 DGIIVFSLSGLFQNVTEKANIIVASALAOIPQKLVMPDGGKPTLGLNTRLYKWIPOND 359
QY 362 LLGHPKTKAFITTHGAMNGIYEAIVHGVPMVGPPIFGDQLDNIAMKKAAGAAVEINFKMT 421
DB 360 LLGHPKTKAFITTHGAMNGIYEAIVHGVPMVGPPIFGDQLDNIAMKKAAGAAVEINFKMT 419
QY 422 SEDLLRALRTVITDSSYKENMRLSRHHDPVYKPLDRAVWIEFWMRHKAQLRSAAH 481
DB 420 SEDLLRALRTVITDSSYKENMRLSRHHDPVYKPLDRAVWIEFWMRHKAQLRSAAH 479
QY 482 DLTWFOHYSIDVIGLITCVATALEFLTKCGLFSCQENKTRKLEKRE 529
DB 480 DLTWFOHYSIDVIGLITCVATALEFLTKCGLFSCQENKTRKLEKRE 527
RESULT 14
AAU7927
ID AAU7927 standard; Protein; 527 AA.
XX
AC AAU7927;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human drug-metabolising enzyme.
XX
DE Human drug-metabolising enzyme; chromosome 4; therapeutic agent;
XX
KW Human; drug-metabolising enzyme; chromosome 4; therapeutic agent;
XX

DR WPI: 2002-066363/09.
 DR N-PSDB: AAD24666.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -

XX
 PS Claim 1a, Page 126-127; 143pp; English.

CC The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-1 protein.

XX
 SQ Sequence 527 AA;

Query Match 68.8%; Score 1924.5; DB 23; Length 527;
 Best Local Similarity 69.9%; Pred. No. 2,2e-179;
 Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKWTSAALLIQLSCYFSSGCGKVLVPTFESHMNIKTIDELVQRGHEVTLASAS 61
 DB 3 SDKSALVFLQLFC-VCGGFCGKVLVPCDMSHMLNVKILIELIVRGHEVTLTHSKP 61
 QY 62 ISFDNPSSTLKFEYVPVSLTTEFEDIKQVYKMAELPKDTFWSYFSQVQEIIMTFND 121
 DB 62 SLIDYRKPSALKFEYVHPQDRTENEIFVDALN--VLPGLSTWOSVIKLDFEVEIRG 119
 QY 122 ILRKFCDIVSNKMLKQLQESRFDVLADAVFPFGLAEHLKIPFVYSLRFSQVAIE 181
 DB 120 TLKMWCESTFYQTLMKLQETNYVMDLPIYPCGDLMABELLAVPVLTLRISVGANME 179
 QY 182 KHSGLLPSPSYVPMVMSLSDQMTFIERVKMITYLVYFEFWQLPDMKKMDQFYSEVLG 241
 DB 180 RSCGKLPAPLSYVPVMTGLTDRMTFLERVKMSMLSVLFHWIODYDYHFWEEFYSKALG 239
 QY 242 RPTTSETMAKADILIRYVNFOPRPHLLPNVEFVGGLHCKPAKPLPKMEEFVQSSGE 301
 DB 240 RPTTCEVVGKAEIMLIRTYWDFEPPQPPNFEFVGGLHCKPAKPLPKMEENFVQSSGE 299
 QY 302 NGVVFSLGSWVNTSEBRANVYASALAKIPKVLMPFQNKPDITGLNTRLYKWIPOND 361
 DB 300 DGIIVFSLGSLFQNTVEEKANIIASALAKIPKVLMPYKGRKPSLTGANTRLYDWIPOND 359
 QY 362 LLGHPKTKAFITHGNGGIYEAIIYGVPMVGPVIFGDQLDNIAMKAKGAIVEINFKTWT 421
 DB 360 LLGHPKTKAFITHGNGGIYEAIIYGVPMVGPVIFGDQLDNIAMKAKGAIVEINFKTWT 419

QY 422 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFWRRHKGAHLSSAAH 481
 DB 420 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFWRRHKGAHLSSAAH 479
 QY 482 DLTWFOHYSIDVIGFLITCVATAIPLFTKCFPLSCQFNKTRIKERE 529
 DB 480 DLTWFOHYSIDVIGFLITCVATAIPLFTKCFPLSCQFNKTRIKERE 527

Search completed: December 5, 2003, 09:48:21
 Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:46:35 ; Search time 21 Seconds

(without alignments)
2422.535 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798

Sequence: 1 MSKMTSALLLIQSCYFSS.....KCFLFSCQKFKTRIKRE 529

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2590.5	92.6	528	2 JN0619	glucuronosyltransf
2	2330	83.3	529	2 A35366	glucuronosyltransf
3	2287	81.7	529	2 J50200	orphan UDP-glucuro
4	2281.5	81.5	528	2 JN0620	UDP-glucuronosyltr
5	2158.5	77.1	530	2 A48633	glucuronosyltransf
6	2086.5	74.6	523	2 S11309	glucuronosyltransf
7	1991.5	71.2	530	2 C47113	glucuronosyltransf
8	1975.5	70.6	529	2 A42233	glucuronosyltransf
9	1938	69.3	531	2 B47113	glucuronosyltransf
10	1930.5	69.0	530	2 S68200	glucuronosyltransf
11	1904.5	68.1	530	2 S00163	glucuronosyltransf
12	1892.5	67.6	530	2 S07390	glucuronosyltransf
13	1853.5	66.2	530	2 A40467	glucuronosyltransf
14	1840.5	65.8	530	2 A36276	glucuronosyltransf
15	1760.5	62.9	527	2 A51089	glucuronosyltransf
16	1760.5	62.8	527	2 A51089	glucuronosyltransf
17	1737.5	60.7	533	2 A35343	glucuronosyltransf
18	1737.5	60.7	533	2 A35343	glucuronosyltransf
19	1737.5	60.7	533	2 A35343	glucuronosyltransf
20	1737.5	60.7	533	2 A35343	glucuronosyltransf
21	1737.5	60.7	533	2 A35343	glucuronosyltransf
22	1737.5	60.7	533	2 A35343	glucuronosyltransf
23	1737.5	60.7	533	2 A35343	glucuronosyltransf
24	1737.5	60.7	533	2 A35343	glucuronosyltransf
25	1737.5	60.7	533	2 A35343	glucuronosyltransf
26	1737.5	60.7	533	2 A35343	glucuronosyltransf
27	1737.5	60.7	533	2 A35343	glucuronosyltransf
28	1737.5	60.7	533	2 A35343	glucuronosyltransf
29	1737.5	60.7	533	2 A35343	glucuronosyltransf

30	490	17.5	534	2 T19944	hypothetical prote
31	489	17.5	485	2 T13694	glucuronosyltransf
32	478.5	17.1	537	2 T21823	hypothetical prote
33	453	16.2	573	2 T27578	hypothetical prote
34	442	15.8	534	2 T19951	hypothetical prote
35	439	15.7	586	2 T19075	hypothetical prote
36	433.5	15.5	502	2 T25263	hypothetical prote
37	428	15.3	949	2 T18591	hypothetical prote
38	427	15.3	530	2 T19365	hypothetical prote
39	423	15.1	745	2 T23893	hypothetical prote
40	419	15.0	534	2 T34455	hypothetical prote
41	410.5	14.7	515	2 T32217	hypothetical prote
42	409	14.6	579	2 H88632	protein F56B3.7 [i
43	404.5	14.5	661	2 T32518	hypothetical prote
44	403	14.4	515	2 S52453	ecdysteroid UDP-gl
45	400	14.3	580	2 T28725	hypothetical prote

ALIGNMENTS

RESULT 1

glucuronosyltransferase (EC 2.4.1.17) 2B-4 precursor - human
JN0619
glucuronosyltransferase (EC 2.4.1.17) 2B-4 precursor - human
N:Alternate names: UDP-glucuronosyltransferase 2B-11
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2000
C:Accession: JN0619; A27878
R:Jin, C.U.; Miners, J.O.; Lillywhite, K.J.; Mackenzie, P.I.
Biochem. Biophys. Res. Commun. 194, 496-503, 1993
A:Title: cDNA cloning and expression of two new members of the human liver UDP-glucurono
A:Reference number: JN0619; MUID:93326164; PMID:8333863
A:Accession: JN0619
A:Molecule type: mRNA
A:Residues: 1-528 <JIN>
A:Cross-references: GB:AF081793; NID:G3426331
A:Experimental source: liver
R:Jackson, M.R.; McCarthy, L.R.; Harding, D.; Wilson, S.; Coughtrie, M.W.H.; Burchell, B
Biochem. J. 242, 581-588, 1987
A:Title: Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA.
A:Reference number: A27878; MUID:87241362; PMID:3109396
A:Accession: A27878
A:Molecule type: mRNA
A:Residues: 1-108, 'F', 110-170, 'RP', 173-381, 'K', 383-384, 'SPR', 388-395, 'F', 397-528 <JAC>
A:Cross-references: GB:Y00317; NID:G37588; PIDN:CAA68415.1; PID:G37589
C:Genetics:
A:Gene: GDB:UGT2B4; UGT2B11
A:Cross-references: GDB:5891331; OMIM:600067
A:Map position: 4q13-4q13
C:Superfamily: glucuronosyltransferase
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-528/Product: glucuronosyltransferase 2B-11 #status predicted <MET>
F:492-509/Domain: transmembrane #status predicted <TM>
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.6% Score 2590.5; DB 2; Length 528;
Best Local Similarity 92.6% Pred. No. 7.2e-192;
Matches 490; Conservative 14; Mismatches 24; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYFSSGCGKVLWPTFSHMNITITIDELVQRGHEVTLASSA 60
DB 1 MSKMTSALLLIQSCYFSSGCGKVLWPTFSHMNITITIDELVQRGHEVTLASSA 60
QY 61 SIFDPNPSPTLKFEEYVPSVLTTFEDIIKQLVKMAELPKOTFMSYFSQVOEIMWTFN 120
DB 61 SIFDPNPSPTLKFEEYVPSVLTTFEDIIKQLVKMAELPKOTFMSYFSQVOEIMWTFN 120
QY 121 DIRKPCQIVSKKMLKMLKQESRPVVLADAVPPGELLAEELKIPVYSIAFSPGYAI 180
DB 121 DIRKPCQIVSKKMLKMLKQESRPVVLADAVPPGELLAEELKIPVYSIAFSPGYAI 180
QY 121 DIRKPCQIVSKKMLKMLKQESRPVVLADAVPPGELLAEELKIPVYSIAFSPGYAI 180
DB 121 DIRKPCQIVSKKMLKMLKQESRPVVLADAVPPGELLAEELKIPVYSIAFSPGYAI 180
QY 181 EKSGGLFPSPVVPVMSLSDQMTFIERVKMIVLYPEFWMFQIFDMKKMDQFSEVL 240
DB 181 EKSGGLFPSPVVPVMSLSDQMTFIERVKMIVLYPEFWMFQIFDMKKMDQFSEVL 240

```
|||||
Db 181 EKHSGLFPSPSYVPMVSELSDMQTFIERVYKMIYVLYFEFVQIPIFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
Db 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
Qy 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Db 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Qy 421 TSBDLRLARLTVITDSSYKENAMRLSRHHDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Db 421 SSTDLNALKVINDPDKENAMKLSRIHQDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
```

RESULT 2

```
A35366
glucuronosyltransferase (EC 2.4.1.17) UDPGth-2 precursor - human
C/Species: Homo sapiens (man)
C/Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 29-Sep-1999
C/Accession: A35366
R/Ritter, J.K.; Sheen, Y.Y.; Owens, I.S.
J. Biol. Chem. 265: 7900-7906, 1990
A/Title: Cloning and expression of human liver UDP-glucuronosyltransferase in COS-1 cell
A/Reference number: A35366; MUID:90243659; PMID:2159463
A/Accession: A35366
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-529 <Rit>
A/Cross-references: GB:J05428; NID:q340079; PIDN:AAA36793.1; PID:q340080
C/Genetics:
A/Gene: GDB:UGT2B7; UGT2B9
A/Cross-references: GDB:5892203; OMIM:600218
A/Map position: 4q13-4q13
C/Superfamily: glucuronosyltransferase
C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein
```

Query Match 83.3%; Score 2330; DB 2; Length 529;

Best Local Similarity 81.1%; Pred. No. 9, 1e-172;

Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```
Qy 1 MSKMTSALLLIQLSCYSSGCGKVLVMPFEFHMNIKITLDELVQRGHEVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYSSGCGKVLVMPFEFHMNIKITLDELVQRGHEVTVLASSA 60
Qy 61 SISPDNPSTLTKFEVYVPSLTKEFEDIKQVLRMAELPKDTFMSYFSQVQIIMTFN 120
Db 61 SISPDNPSTLTKFEVYVPSLTKEFEDIKQVLRMAELPKDTFMSYFSQVQIIMTFN 120
Qy 121 DLRKFKCDIYSNKKLKKQESRPDVYLAADVFPFGEILAEELKIPVYSLRFSGYAI 180
Db 121 DLRKFKCDIYSNKKLKKQESRPDVYLAADVFPFGEILAEELKIPVYSLRFSGYAI 180
Qy 181 EKHSGLFPSPSYVPMVSELSDMQTFIERVYKMIYVLYFEFVQIPIFMKMDQFYSVL 240
Db 181 EKHSGLFPSPSYVPMVSELSDMQTFIERVYKMIYVLYFEFVQIPIFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
Db 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
```

```
Qy 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Db 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Qy 421 TSBDLRLARLTVITDSSYKENAMRLSRHHDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Db 421 SSTDLNALKVINDPDKENAMKLSRIHQDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
```

RESULT 3

```
JE0200
orphan UDP-glucuronosyltransferase (EC 2.4.-.-) - human
N/Alternate names: UGT2B11
C/Species: Homo sapiens (man)
C/Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C/Accession: JE0200
R/Beaulieu, M.; Levesque, E.W.; Hum, D.; Belanger, A.
Biochem. Biophys. Res. Commun. 248: 44-50, 1998
A/Title: Isolation and characterization of a human orphan UDP-glucuronosyltransferase
A/Reference number: JE0200; MUID:98340847; PMID:9675083
A/Accession: JE0200
A/Molecule type: mRNA
A/Residues: 1-529 <BBA>
A/Cross-references: GB:AF016492
A/Comment: This enzyme catalyzes the transfer of glucuronic acid to a wide range of exog
C/Superfamily: glucuronosyltransferase
C/Keywords: glycoprotein; glycosyltransferase
F:493-509/domain: membrane-anchoring #status predicted <MC>
F:315/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

Query Match 81.7%; Score 2287; DB 2; Length 529;

Best Local Similarity 80.0%; Pred. No. 1, 9e-168;

Matches 423; Conservative 43; Mismatches 63; Indels 0; Gaps 0;

```
Qy 1 MSKMTSALLLIQLSCYSSGCGKVLVMPFEFHMNIKITLDELVQRGHEVTVLASSA 60
Db 1 MTLKMTSVLLIHLISCFSSGCGKVLVMAEYSHMMNKITLDELVQRGHEVTVLASSA 60
Qy 61 SISPDNPSTLTKFEVYVPSLTKEFEDIKQVLRMAELPKDTFMSYFSQVQIIMTFN 120
Db 61 SILPDNPASTLTKFEVYVPSLTKEFENIMQVLRMSDIPKDSFMYLFSQVQIIMTFN 120
Qy 121 DLRKFKCDIYSNKKLKKQESRPDVYLAADVFPFGEILAEELKIPVYSLRFSGYAI 180
Db 121 DIFNPKCDIVSNKKKKQESRPDIYFADAVFPFGEILAEELKIPVYSLRFSGYAI 180
Qy 181 EKHSGLFPSPSYVPMVSELSDMQTFIERVYKMIYVLYFEFVQIPIFMKMDQFYSVL 240
Db 181 ERHSGGLFPSPSYVPMVSELSDMQTFIERVYKMIYVLYFEFVQIPIFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDQFPPHPLPNVEFGVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
Db 301 ENGIVVPSLGSVNTSEERANVIASALAKIPOKVLMPFDGNKPTDGLNRLYKWIPON 360
Qy 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Db 361 DLGHPKTKAFITGGNNGIYEALYHGVPMGVPIFGDQDNLNIAHMAKGAABVEINFTM 420
Qy 421 TSBDLRLARLTVITDSSYKENAMRLSRHHDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Db 421 SSTDLNALKVINDPDKENAMKLSRIHQDPVPLDRAVFWIEFVRRHKGAKHLRSAA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFSCKFNKTRKIEKRE 529
```

Db 481 HDLTFWFOHSLDVIIGFLACVATVFIITKFCFCFMKFAKGGKKRD 529

RESULT 4

JN0620

UDP-glucuronosyltransferase (EC 2.4.1.-) 2B-10 precursor - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C/Accession: JN0620

R/In: C.J.; Miners, J.O.; Lillywhite, K.J.; Mackenzie, P.I.

Biochem. Biophys. Res. Commun. 194, 496-503, 1993

A/Title: cDNA cloning and expression of two new members of the human liver UDP-glucuronosyltransferase family: glucuronosyltransferase; hexosyltransferase; transmembrane protein

A/Reference number: JN0619; MUID:93326164; PMID:8333863

A/Accession: JN0620

A/Molecule type: mRNA

A/Residues: 1-528 <JN>

A/Cross-references: GB:X63359; NID:G516149; PIDN:CAA44961.1; PID:G516150

A/Experimental source: liver

C/Superfamily: glucuronosyltransferase

C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-528/Product: UDP-glucuronosyltransferase 2B-10 #status predicted <MET>

F:491-508/Domain: transmembrane #status predicted <TM>

F:66,314,481/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 81.5%; Score 2281.5; DB 2; Length 528;

Best Local Similarity 80.5%; Pred. No. 5e-168;

Matches 426; Conservative 44; Mismatches 56; Indels 1; Gaps 1;

Qy 1 MSMTSALLLIQLSCYFSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60
Db 1 MALKMTT-VLLIQSLFYSSGCGKLVMAEYSLMMNMKITILKELVQGHVTLASSA 59

Qy 61 SISDPNSPSTLKEVYVVSITKTEFEDIIKQLVKRAELPKDTFMSYFSQVQELMTWF 120
Db 60 SILDPNDSSITLKEVYVPTSLTKTEFENIIMQLVKRSLEIQDPTFMSYFSQVQELMTWF 119

Qy 121 DILKRCFDIVSNKMKLQESRPDYVLADAVFPFGELLAELKIPVYSLRSPGYAI 180
Db 120 DIINFCQDVSNKMKLQESRPDYVLADAVFPFGELLAELNIPVYSHSFGYSF 179

Qy 181 EKHSGLLPPSPYVPMVSELSDDMTFIERVKMNIYVLYFEFMPQIFDMKKMDQFSEVL 240
Db 180 ERHSGGFPPSPYVPMVSKLSDMTFIERVKMNIYVLYFPFQIFDMKKMDQFSEVL 239

Qy 241 GRPTTLESETMAKADIWLRNWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQSS 300
Db 240 GRPTTLESETMAKADIWLRNWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQSS 299

Qy 301 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDPGNKPDGLNTRLYKWIPO 360
Db 300 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDPGNKPDGLNTRLYKWIPO 359

Qy 361 DLLGHPKTAFTTHGGMNGIYEALYHGVPMVGPVIFGDDLDNIAMKAKGAIVEINFTM 420
Db 360 DLLGHPKTAFTTHGGMNGIYEALYHGVPMVGPVIFGDDLDNIAMKAKGAIVVDNTM 419

Qy 421 TSEDLRALRTVITDSSYKENAMRLSRHHDDQVYKPLDRAVFWIEFVNRHKGALRSAA 480
Db 420 SSTDLNALKTVINDPSYKENIMKLSRIQHDQVYKPLDRAVFWIEFVNRHKGALRSAA 479

Qy 481 HDLTFWFOHSLDVIIGFLACVATVFIITKFCFCFMKFAKGGKKRD 529
Db 480 HDLTFWFOHSLDVIIGFLACVATVFIITKFCFCFMKFAKGGKKRD 528

RESULT 5

A48633

glucuronosyltransferase (EC 2.4.1.17) precursor - human
N/Alternate names: dihydrotestosterone/androstenediol UDP-glucuronosyltransferase isoform
C/Species: Homo sapiens (man)
C/Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 29-Sep-1999
C/Accession: A48633; I38559

R/Chen, F.; Ritter, J.K.; Wang, M.G.; McBride, O.W.; Lubet, R.A.; Owens, I.S.

Biochemistry 32, 10648-10657, 1993

A/Title: Characterization of a cloned human dihydrotestosterone/androstenediol UDP-glucuronosyltransferase

A/Reference number: A48633; MUID:94002056; PMID:8399210

A/Accession: A48633

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-530 <RES>

A/Experimental source: liver

A/Note: sequence extracted from NCBI backbone (NCBI:138786)

R/Green, M.D.; Oturu, E.M.; Tepily, T.R.

Drug Metab. Dispos. 22, 799-805, 1994

A/Title: Stable expression of a human liver UDP-glucuronosyltransferase (UGT2B15) with a

A/Reference number: I38559; MUID:95136867; PMID:7835232

A/Accession: I38559

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-530 <RES>

A/Cross-references: EMBL:U08854; NID:9475758; PIDN:AAC50077.1; PID:9475759

C/Genetics:

A/Gene: GDB:UGT2B15; UGT2B8

A/Cross-references: GDB:5892418; OMIM:600219

A/Map position: 4q13-4q13

C/Superfamily: glucuronosyltransferase

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 77.1%; Score 2158.5; DB 2; Length 530;

Best Local Similarity 74.5%; Pred. No. 1.6e-158;

Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

Qy 1 MSMTSALLLIQLSCYFSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60
Db 1 MSMTSALLLIQLSCYFSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60

Qy 61 SISDPNSPSTLKEVYVVSITKTEFEDIIKQLVKRAELPKDTFMSYFSQVQELMTWF 119
Db 60 SILDPNDSSITLKEVYVPTSLTKTEFENIIMQLVKRSLEIQDPTFMSYFSQVQELMTWF 120

Qy 120 NDILKRCFDIVSNKMKLQESRPDYVLADAVFPFGELLAELKIPVYSLRSPGYAI 179
Db 121 YDYSNKLCKDAVSNKMKLQESRPDYVLADAVFPFGELLAELKIPVYSLRSPGYAI 180

Qy 180 IEKHSGLLPPSPYVPMVSELSDDMTFIERVKMNIYVLYFEFMPQIFDMKKMDQFSEV 239
Db 181 FEKNGGFLPPSPYVPMVSKLSDMTFIERVKMNIYVLYFPFQIFDMKKMDQFSEV 240

Qy 240 GRPTTLESETMAKADIWLRNWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQSS 299
Db 241 LGRTTLESETMAKADIWLRNWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQSS 300

Qy 300 GENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDPGNKPDGLNTRLYKWIPO 359
Db 301 GENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDPGNKPDGLNTRLYKWIPO 360

Qy 360 NDILGHPKTAFTTHGGMNGIYEALYHGVPMVGPVIFGDDLDNIAMKAKGAIVEINFTM 419
Db 361 NDILGHPKTAFTTHGGMNGIYEALYHGVPMVGPVIFGDDLDNIAMKAKGAIVVDNTM 420

Qy 420 MTSDDLRLALRTVITDSSYKENAMRLSRHHDDQVYKPLDRAVFWIEFVNRHKGALRSAA 479
Db 421 MSSRDLNALKTVINDPSYKENIMKLSRIQHDQVYKPLDRAVFWIEFVNRHKGALRSAA 480

Qy 480 HDLTFWFOHSLDVIIGFLACVATVFIITKFCFCFMKFAKGGKKRD 529
Db 481 HDLTFWFOHSLDVIIGFLACVATVFIITKFCFCFMKFAKGGKKRD 530

RESULT 6

S11309

glucuronosyltransferase (EC 2.4.1.17) - human
N/Alternate names: estradiol UDPglucuronosyltransferase
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999

C:Accession: S11309
 R:Coffman, B.L.; Tephly, T.R.; Irshaid, Y.M.; Green, M.D.; Smith, C.; Jackson, M.R.; Moc Arch. Biochem. Biophys. 261, 170-175, 1990
 A:Title: Characterization and primary sequence of a human hepatic microsomal estriol UDP A:Reference number: S11309; MUID:90343358; PMID:2116769
 A:Accession: S11309
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-523 <COB>
 A:Cross-references: GB:U06641; NID:g458398; PIDN:AAA83406.1; PID:g458399
 A:Note: the authors translated the codon AGG for residue 412 as Thr and ACC for residue C:Superfamily: glucuronosyltransferase
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 74.6%; Score 2086.5; DB 2; Length 523;
 Best Local Similarity 73.9%; Pred. No. 5.5e-153;
 Matches 385; Conservative 57; Mismatches 78; Indels 1; Gaps 1;

Qy 10 LLIQLSCYFSSGSCGKVLVWPTFESHMNNIKITIDELVQGHVTVLSSASISFDPSP 69
 Db 3 LLIQLSCYFSSGSCGKVLVWPTFESHMNNIKITIDELVQGHVTVLSSASISFDPSP 62
 Qy 70 STLFVYVPSVLTKEFEDIKQVLKVA-ELPKDTPMSYFSQVQELMTNDILKFCX 128
 Db 63 SAIKLEVYPTSLTKNYLSDSLKIDRWIYGVSKTFFMSYFSQVQELCMAYDYSNKLCX 122
 Qy 129 DIVSKTKMLKQESRFDVLADAFPPGELLAELKTPVYSLSRFSFGYAIKHSGLL 188
 Db 123 DAVLKKMLKMLKQESRFDVLPDVPFPGCELLSELINLPVYRSRISRGYTIKNGGGL 182
 Qy 189 FPPSVVPMSELSDQMTFIERVKMIYLYFEFQIYFDKMKMQDFSEVLGRFTTSE 248
 Db 183 FPPSVVPMSELSDQMTFIERVKMIYLYFEFQIYFDKMKMQDFSEVLGRFTTSE 242
 Qy 249 TMAKADIWLRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSSGNGVVS 308
 Db 243 TMGKAEMLIRTYWDFEPPRPPLPVVDVGGHCKPAKPLPKLEEFVQSSGNGVVS 302
 Qy 309 LGSVNSNSERANVIALAKIPQKVLMPDGNKPDLTGLNTRLYKMTIPONDILGHKT 368
 Db 303 LGSVNSNSERANVIALAKIPQKVLMPDGNKPDLTGLNTRLYKMTIPONDILGHKT 362
 Qy 369 KAFITGGMNGIYEAIYHGVPMVGPPIFGDLDNIJAHKAKAAGVINEFKTMSSEDLRA 428
 Db 363 KAFITGGMNGIYEAIYHGVPMVGPPIFGDLDNIJAHKAKAAGVINEFKTMSSEDLRA 422
 Qy 429 LRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVRRHKGAKHLRSAADLTWFOH 488
 Db 423 LKSVINDVYKENAMKLSRIHHDQMKPLDRAVFWIEFVRRHKGAKHLRSAADLTWFOH 482
 Qy 489 YSIDVIGFLTCVATAIPLFTKCFPLFSQCKFKTKIKIEKE 529
 Db 483 HSLDVIAFLWAVATVFIITKFCFLCFRKLAKKKKKRD 523

RESULT 7
 C47113
 Glucuronosyltransferase (EC 2.4.1.17) UGT2B14 precursor - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
 A:Accession: C47113
 R:Tukey, R.H.; Pendurthi, U.R.; Nguyen, N.T.; Green, M.D.; Tephly, T.R.
 J. Biol. Chem. 268, 15260-15266, 1993
 A:Title: Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs.
 A:Reference number: A47113; MUID:9315511; PMID:8325897
 A:Accession: C47113
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-530 <TVK>
 A:Cross-references: GB:L01082; NID:g165798; PIDN:AAA18021.1; PID:g165799
 C:Superfamily: glucuronosyltransferase
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 71.2%; Score 1991.5; DB 2; Length 530;
 Best Local Similarity 67.9%; Pred. No. 1.2e-145;
 Matches 360; Conservative 79; Mismatches 90; Indels 1; Gaps 1;

Qy 1 MSKMTBA-LLIQLSCYFSSGSCGKVLVWPTFESHMNNIKITIDELVQGHVTVLSS 59
 Db 1 MSKMTBA-LLIQLSCYFSSGSCGKVLVWPTFESHMNNIKITIDELVQGHVTVLSS 59
 Qy 60 ASISFDNSPSTLKEEYVPSVLTKEFEDIKQVLKVAELPKDTPMSYFSQVQELMTND 119
 Db 61 ASISFDNSPSTLKEEYVPSVLTKEFEDIKQVLKVAELPKDTPMSYFSQVQELMTND 120
 Qy 120 NDILRKECDIVSKTKMLKQESRFDVLADAFPPGELLAELKTPVYSLSRFSFGYAI 179
 Db 121 SDSCGNCKEYVPMKTLMTKQESRFDVLADAFPPGELLAELKTPVYSLSRFSFGYAI 180
 Qy 180 IEKHSGLLFPSPVPMSELSDQMTFIERVKMIYLYFEFQIYFDKMKMQDFSEV 239
 Db 181 MEKYSGLSVPPSYVPIILSDLSGKMTFMRVNNMLCMLYDFWFMNKKRMDQFYSEV 240
 Qy 240 LGRFTTSETMAKADIWLRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSS 299
 Db 241 LGRFTTSETMAKADIWLRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSS 300
 Qy 300 GENGVVFSLGSVNSNSERANVIALAKIPQKVLMPDGNKPDLTGLNTRLYKMTIPQ 359
 Db 301 GENGVVFSLGSVNSNSERANVIALAKIPQKVLMPDGNKPDLTGLNTRLYKMTIPQ 360
 Qy 360 NDILGHPRYKAFITGGMNGIYEAIYHGVPMVGPPIFGDLDNIJAHKAKAAGVINEFKT 419
 Db 361 NDILGHPRYKAFITGGMNGIYEAIYHGVPMVGPPIFGDLDNIJAHKAKAAGVINEFKT 420
 Qy 420 MTSSEDLNALKTVINDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVRRHKGAKHLRSA 479
 Db 421 MTSSEDLNALKTVINDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVRRHKGAKHLRSA 480
 Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIPLFTKCFPLFSQCKFKTKIKIEKE 529
 Db 481 AHDLTWFOHYSIDVIGFLTCVATAIPLFTKCFPLFSQCKFKTKIKIEKE 530

RESULT 8
 A42233
 Glucuronosyltransferase (EC 2.4.1.17) 2 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Jun-1992 #sequence_revision 18-Sep-1992 #text_change 29-Sep-1999
 A:Accession: A42233; A24324
 R:Mackenzie, P.I.; Rodbourn, L.
 J. Biol. Chem. 265, 11328-11332, 1990
 A:Title: Organization of the rat UDP-glucuronosyltransferase, UDPGT-2, gene and character.
 A:Reference number: A42233; MUID:90293083; PMID:2113533
 A:Accession: A42233
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-529 <MAC>
 A:Cross-references: GB:J05482
 A:Note: the authors translated the codon GTA for residue 57 as B, and GNC for residue 11
 R:Mackenzie, P.I.
 J. Biol. Chem. 261, 6119-6125, 1986
 A:Title: Rat liver UDP-glucuronosyltransferase. Sequence and expression of a cDNA encodi
 A:Reference number: A24324; MUID:86196018; PMID:3084479
 A:Accession: A24324
 A:Molecule type: mRNA
 A:Residues: 1-407, 'V', 409-529 <MA2>
 A:Cross-references: GB:M3506; NID:g207580; PIDN:AAA42313.1; PID:g207581
 A:Experimental source: liver
 C:Superfamily: glucuronosyltransferase
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 70.6%; Score 1975.5; DB 2; Length 529;
 Best Local Similarity 68.4%; Pred. No. 2e-144;
 Matches 361; Conservative 68; Mismatches 98; Indels 1; Gaps 1;

```

Qy 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
Db 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
Qy 61 SISDPSPSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 119
Db 61 SISDPSPSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 119
Qy 120 NDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGYA 179
Db 120 NDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGYA 179
Qy 121 SDVVENICKALIMKSLMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGYR 180
Db 121 SDVVENICKALIMKSLMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGYR 180
Qy 180 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 239
Db 180 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 239
Qy 181 CEKSGGLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 240
Db 181 CEKSGGLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 240
Qy 240 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 299
Db 240 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 299
Qy 241 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 300
Db 241 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 300
Qy 300 GENGWVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 359
Db 300 GENGWVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 359
Qy 301 GENGWVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 360
Db 301 GENGWVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 360
Qy 360 NDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 419
Db 360 NDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 419
Qy 361 NDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 420
Db 361 NDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 420
Qy 420 MTSEDLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 479
Db 420 MTSEDLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 479
Qy 421 LSTGLLTALAKIVWMDPSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRST 480
Db 421 LSTGLLTALAKIVWMDPSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRST 480
Qy 480 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 527
Db 480 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 527
Qy 481 LHDLSWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 528
Db 481 LHDLSWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 528

```

RESULT 9

```

B47113
glucuronosyltransferase (EC 2.4.1.17) UGT2B13 precursor - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
R/Key: B47113
R/Title: Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs
J. Biol. Chem. 268, 15260-15266, 1993
A/Accession: B47113
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-531 <TRK>
A/Cross-references: GB:L01081; NID:g165796; PIDN:AAA18020.1; PID:g165797
C/Superfamily: glucuronosyltransferase
C/Keywords: glucosyltransferase; hexosyltransferase; transmembrane protein

```

```

Query Match 69.3%; Score 1938; DB 2; Length 531;
Best Local Similarity 68.7%; Pred. No. 1.6e-141;
Matches 365; Conservative 62; Mismatches 102; Indels 2; Gaps 2;
Qy 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 59
Db 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
Qy 60 ASISFDNPSSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 118
Db 60 ASISFDNPSSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 118
Qy 61 ASISFDNPSSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 120
Db 61 ASISFDNPSSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 120
Qy 119 FNDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGY 178
Db 119 FNDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGY 180
Qy 121 YSDICEDICKEVILMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGY 180
Db 121 YSDICEDICKEVILMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGY 180
Qy 179 AIEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSE 238
Db 179 AIEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSE 238

```

```

Db 181 MLOKGGGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSE 240
Qy 239 VLRGPTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 298
Db 239 VLRGPTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 298
Qy 241 VLRGPTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 300
Db 241 VLRGPTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 300
Qy 299 SGENGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 358
Db 299 SGENGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 358
Qy 301 SGENGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 360
Db 301 SGENGVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPDGNKPDILGINTRLYKMIPO 360
Qy 359 QNDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 418
Db 359 QNDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 418
Qy 361 QNDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 420
Db 361 QNDILGHPKTAFTTHGNGNGIYEAIVHGVPMVGPPIFGDQDNIAMKAKAAVEINFKT 420
Qy 419 TMSADLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 478
Db 419 TMSADLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 478
Qy 421 TMSADLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 480
Db 421 TMSADLRLPRTVITDSSYKENAMRLSRHHQDPVKPLDRAVFIEMVRHKGAKHLRSA 480
Qy 479 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 529
Db 479 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 529
Qy 481 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 531
Db 481 AHDLTWFOHYSIDVIGFLTLCVATAIFLTCKCFELFSCQKFNKTRKIEKE 531

```

RESULT 10

```

S68200
glucuronosyltransferase (EC 2.4.1.17) precursor - rat
N/Alternate names: UDP-glucuronosyltransferase
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000
A/Accession: S68200; A61266
R/Key: S68200; A61266
R/Title: Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase
Arch. Biochem. Biophys. 322, 460-468, 1995
A/Accession: S68200; MIMD:96032669; PMID:7514722
A/Reference number: S68200; MIMD:96032669; PMID:7514722
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-530 <GRE>
A/Cross-references: EMBL:U06273; NID:g458394; PIDN:AAA83404.1; PID:g458395
R/Scyzymski, P.; Green, M.; Fuld, J.; Coffman, B.; Tephly, T.
Mol. Pharmacol. 40, 80-84, 1991
A/Title: Purification and properties of a rat liver phenobarbital-inducible 4-hydroxybiphenyl-2-yl-23/Domain: signal sequence #status predicted <SIG>
F:1-23/Product: glucuronosyltransferase #status predicted <MAT>

```

```

Query Match 69.0%; Score 1930.5; DB 2; Length 530;
Best Local Similarity 66.8%; Pred. No. 6e-141;
Matches 354; Conservative 73; Mismatches 102; Indels 1; Gaps 1;
Qy 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
Db 1 MSKMTSALLLIQUSCFSSGSGCKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
Qy 61 SISDPSPSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 119
Db 61 SISDPSPSTLKEEYVPSVLTKEFEEDIKQVYKMA-ELPKOTFMSYSQVQEIWTF 119
Qy 61 FVLDPKETSDLKFTVPTSPSSSHDLENFTFRVANNVYTLPEPDCTSLYLYQDTIDEX 120
Db 61 FVLDPKETSDLKFTVPTSPSSSHDLENFTFRVANNVYTLPEPDCTSLYLYQDTIDEX 120
Qy 120 NDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGYA 179
Db 120 NDILRKFCQDIVSKKLMKKLQESRPDVVLADAVFPGEELAEELKIPVYSLRFSFGYA 179
Qy 121 SDVVENICKALIMKSLMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGYR 180
Db 121 SDVVENICKALIMKSLMKLQSSQFDVILADAVGCEELAEELKIPVYSLRFSFGYR 180
Qy 180 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 239
Db 180 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 239
Qy 181 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 240
Db 181 IEKHSGLLPSPSYVPMVSELSQMTFIERVKMIVYLFEFWMFOIFDMKKMDQFSEV 240
Qy 240 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 299
Db 240 LGRTTLETMAKADILIRNYWPOPHPLPNVEFVGGIHCCKPAKPLPKMEEFVQSS 299

```


RESULT 13
A40467
glucuronosyltransferase (EC 2.4.1.17) precursor - rat
N:Alternate names: UDP-glucuronosyltransferase isoform 53K
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Feb-1992 #sequence revision 07-Feb-1992 #text_change 17-Nov-2000
C:Accession: A40467, A23520; S59627; A26064; I55247
R:Hague, S.J.; Petersen, D.D.; Nebert, D.W.; Mackenzie, P.I.
DNA Cell Biol. 10, 515-524, 1991
A:Title: Isolation, sequence, and developmental expression of rat UGT2B2: the gene encoding
A:Reference number: A40467; MUID:91369480; PMID:1909872

A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-530 <HAO>
R:Jackson, M.R.; Burchell, B.
Nucleic Acids Res. 14, 779-795, 1986
A:Title: The full length coding sequence of rat liver androsterone UDP-glucuronyltransferase
Reference number: A23520; MUID:86120371; PMID:3003696

A:Residues: 31-158, 'E', 160-285, 'S', 287-350, 'I', 352-362, 'I', 364-430, 'E', 432-530 <MAC>
A:Cross-references: GB:X03478; NID:g57452; PIDD:CAA7198.1; PID:g57453
A:Note: the authors translated the codon ATT for residue 321 as Asn
Yamashita, A.; Watanabe, M.; Tonegawa, T.; Sugitara, T.; Waku, K.
Biochem. J. 312, 301-308, 1995
A:Title: Acyl-CoA binding and acylation of UDP-glucuronosyltransferase isoforms of rat I
A:Reference number: S59626; MUID:96077159; PMID:7492328
A:Accession: S59627
A:Molecule type: protein
A:Residues: 24-44 <YAM>
R:Mackenzie, P.I.
J. Biol. Chem. 261, 14112-14117, 1986
A:Title: Rat liver UDP-glucuronosyltransferase. cDNA sequence and expression of a form g
A:Reference number: A26064; MUID:87033594; PMID:2429951
A:Accession: A26064
A:Molecule type: mRNA
A:Residues: 1-430, 'E', 432-530 <MAC>
A:Cross-references: GB:U02589; NID:g207582; PIDD:AAA4214.1; PID:g207583
A:Experimental source: hepatic
A:Superfamily: glucuronosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein
E:1-23/Domain: signal sequence #status predicted <SIG>
F:24-530/Product: glucuronosyltransferase #status experimental <MAT>

	Query Match	Score	1853.5	DB 2	Length	530
	Best Local Similarity	64.7%	Pred. No.	5.3e-135		
	Matches	343	Conservative	71	Indels	1
					Gaps	1
Qy	1	MSMKVTSLLDILQSLCSFSSGCGKVLWPTFTFSMMNNIKTILDELVDGRHEVYTLASAA	60			
Db	1	MPRRKISLFLQIQTSCFRSGHGCVLWPMDFSMNNIKTILDELVDGRHEVYTLAKPSA	60			

RESULT 14
A36276
glucuronosyltransferase (EC 2.4.1.17) 5 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Jan-1991 #sequence_revision 13-Jan-1993 #text_change 29-Sep-1999
C:Accession: A36276
R:Mackenzie, P.I.
J. Biol. Chem. 265, 8699-8703, 1990
A:Title: The cDNA sequence and expression of a variant 17beta-hydroxysteroid UDP-glucuron
A:Reference number: A36276; MUID:90256795; PMID:1692835
A:Accession: A36276
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-530 <MAC>
A:Cross-references: GB:M33746; GB:J05440; NID:G207570; PIDN:AAA03217.1; PID:G207572
A:Note: The authors translated the codon GAG for residue 530 as Lys
A:Superfamily: Glucuronosyltransferase

Query Match	65.8%	Score 1840.5	DB 2	Length 530
Best Local Similarity	64.3%	Pred. No. 5.3e-134		
Matches 341	Conservative 73	Mismatches 115	Indels 1	Gaps 1

QY MSKRWISALLLQLOSCYFSSSGSGVLAWPLPESFHMNIKILIDELVORGHEVTLVMSA 60

Db 1 MPEKWIIFALLLLQISPCRLSACGKGVLAWPEMFHMNIKITILDELVORGHEVTLVAKSA 60

QY 61 SISFDPNSSPTLKEFVYVSVLTKTEPEEDIKQVLKRW-ELPKOTFWSYFSCVQVEIMWTF 115

Db 61 YYVLDPKKSPPDLKEFFPTSPVSKDELKEXFIKLADAMVTELRDPTCLSFSLQNMNDEF 120

QY 120 NDLIRKFCDIYSNKKMLKQOESRFDVVLADAVFPBELLAELLIKIPVYSLAFSGYA 175

Db 121 SDYVLSVCDAVSNKQVLAKQOESKFDYLLSPVAAAGELLAEVLAHIFPLVLSLAPSGHK 180

QY 180 IEKHSGLLFPSPYVYVVMSELSDMTFLERKQNIIVYLVEFPMFOIFDMKKMQFOFSEV 235

Db 181 IEKSSGRFLUPSPYVYVILSGQMTFLDRKQNIICMLVDFPMHFMFNACMQMPFTEI 240

Qy 240 LGRPTLSETMAKADIWILIRNYWDFQPHPLPVEVFGIACPKAPKLPKMESEFVOSS 299
Db 241 LGRPTTLAETNGKXEMWILIRSYWDFEPHPLPNDVYIGIACPKAPKLPDIDDFVOSS 300
Qy 300 GENGVVFSLSGWSNTESEEPANVASALAKIPOKVLRFPGDKPDTLGLNTRLYKMIPO 359
Db 301 GEHGVVFSLSGWSNTESEEPANVASALAKIPOKVLRFPGDKIPATIGPNTRYKMLPO 360
Qy 360 NDLLGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDOLNIAHMKAKGAVEINFKT 419
Db 361 NDLLGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDOLNIAHMKAKGAVALNIRT 420
Qy 420 MTEEDLRLALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 479
Db 421 MSKSDLFWALKEVINNPYKXNAMLSITHHDQPKPLDKALFWIEYVRRHKGAKHLRPL 480
Qy 480 AHDLTWFQHSIDVIGFLITCVATAIFLFTKCFLFCCKFNKTRKIEKRE 529
Db 481 GHNLPWYQHSIDVIGFLITCVATAIFLFTKCFLFCCKFNKTRKIEKRE 530

RESULT 15

S15089
glucuronosyltransferase (EC 2.4.1.17) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S15089
R:Lazard, D.; Zupko, K.; Portia, Y.; Neff, P.; Lazarovite, J.; Horn, S.; Khen, M.; Lancet, Nature 349, 790-793, 1991
A:Title: Odorant signal termination by olfactory UDP glucuronosyl transferase.
A:Reference number: S15089; MUID:91156050; PMID:1900353
A:Accession: S15089
A:Molecule type: mRNA
A:Residues: 1-527 <LAZ>
A:Cross-reference: GB:X57565; NID:957762; PIDN:CAA40797.1; PID:93980217
C:Superfamily: glucuronosyltransferase
C:Keywords: glycosyltransferase, hexosyltransferase

Query Match 62.9%; Score 1760.5; DB 2; Length 527;

Best Local Similarity 62.0%; Pred. No. 7.8e-128; Indels 9; Gaps 3;

Matches 327; Conservative 77; Mismatches 114; Indels 9; Gaps 3;

Qy 5 WTSALLLIQLSCYFSSGSGKVLVWPTFSHWNKIKTILDELVGRHEVTVLASSASISF 64
Db 8 WSLQSLSLGMSL-----GNNVLWPMGSHWLVNKKIIDELRKHNVTVLVASGALFI 61
Qy 65 DPNSPSTLKFEYVYSLTKTEFEDIIKQIVKMAE-LPKDTFWSYFSQVOEIMTFENDI 122
Db 62 TFSVSPSLTFEYVYSLTKTEFEDIIKQIVKMAE-LPKDTFWSYFSQVOEIMTFENDI 121
Qy 123 LRKFKDIVSNKKMKKLOESRFVDVLDAVFPGEALLKIPVVSILRSPGATLEK 182
Db 122 SRGIDGVALKNEKMTKLRGKFEVLVLSDPVPCGDIVALKIGIPFISLRSPASTVEK 181
Qy 183 HSGGLLPPPSYVPMVSEISDQMTFIERKNIIVLYFEFQIFDMKKMDQFYSVLGR 242
Db 182 HCGKVPFSPSYVPMVSEISDQMTFIERKNIIVLYFEFQIFDMKKMDQFYSVLGR 240
Qy 243 PTTLSETMAKADIWILIRNYWDFQPHPLPVEVFGIACPKAPKLPKMESEFVOSSGEN 302
Db 241 PTTLSETMAKADIWILIRNYWDFQPHPLPVEVFGIACPKAPKLPKMESEFVOSSGEN 300
Qy 303 GVVVFSLSGWSNTESEEPANVASALAKIPOKVLRFPGDKPDTLGLNTRLYKMIPO 362
Db 301 GVVVFSLSGWSNTESEEPANVASALAKIPOKVLRFPGDKPDTLGLNTRLYKMIPO 360
Qy 363 LGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDOLNIAHMKAKGAVEINFKTMTS 422
Db 361 LGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDOLNIAHMKAKGAVEINFKTMTS 420
Qy 423 EDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 482
Db 421 ADLLSAVAVINEPFYKXNAMLSITHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 480

Qy 483 LTFWQHSIDVIGFLITCVATAIFLFTKCFLFCCKFNKTRKIEKRE 529
Db 481 LSWFQHSIDVIGFLITCVATAIFLFTKCFLFCCKFNKTRKIEKRE 527

Search completed: December 5, 2003, 09:50:16
Job time: 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:46:05 ; Search time 39 Seconds
(without alignments)
3500.252 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSKMTSALLILQSCYFSS.....KCFLFSCQKFKTKIEKRE 529

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeoph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2368.5	84.6	528	6 Q8WN97	Q8WN97 macaca fasc
2	2311	82.6	529	6 Q97951	Q97951 macaca fasc
3	2291	81.9	529	6 Q9GLD9	Q9GLD9 macaca mula
4	2289	81.8	529	6 Q9GLE0	Q9GLE0 macaca mula
5	2261	80.8	529	6 Q9TSL6	Q9TSL6 macaca fasc
6	1999.5	71.5	529	11 Q8R084	Q8R084 mus musculu
7	1993.5	71.2	532	11 Q8K154	Q8K154 mus musculu
8	1970.5	70.4	528	11 Q8VIF9	Q8VIF9 cavia porce
9	1908.5	68.2	530	11 Q8K169	Q8K169 mus musculu
10	1895.5	67.7	530	11 Q91WH2	Q91WH2 mus musculu
11	1895	67.7	529	11 Q8BJU9	Q8BJU9 mus musculu
12	1878	67.1	529	11 Q8VIF8	Q8VIF8 cavia porce
13	1854.5	66.3	530	11 Q8VCN3	Q8VCN3 mus musculu
14	1802	64.4	530	11 Q9R110	Q9R110 cavia porce
15	1795.5	64.2	527	4 Q9Y4X1	Q9Y4X1 homo sapien
16	1780.5	63.6	527	11 Q9ESB4	Q9ESB4 mus musculu

17	1747.5	62.5	534	11 Q8BWQ1	Q8BWQ1 mus musculu
18	1739.5	62.2	534	11 Q8R129	Q8R129 mus musculu
19	1737.5	62.1	534	11 Q9DB11	Q9DB11 mus musculu
20	1721	61.5	449	4 Q9H6S4	Q9H6S4 mus sapien
21	1204.5	43.0	531	11 Q8VD45	Q8VD45 mus sapien
22	1192	42.6	533	6 Q46549	Q46549 felis silve
23	1144.5	40.9	533	6 Q9TSL7	Q9TSL7 macaca fasc
24	1144.5	40.9	533	6 Q9SKM4	Q9SKM4 macaca mula
25	1138.5	40.7	533	6 Q46423	Q46423 felis silve
26	1135.5	40.6	533	6 Q9PDG7	Q9PDG7 macaca mula
27	1132.5	40.5	533	6 Q46548	Q46548 felis silve
28	1121.5	40.1	530	6 Q9TSM0	Q9TSM0 macaca fasc
29	1120.5	40.0	530	4 Q9HAW9	Q9HAW9 homo sapien
30	1120.5	40.0	530	4 Q60656	Q60656 homo sapien
31	1116.5	39.9	530	4 Q8TEX4	Q8TEX4 homo sapien
32	1114.5	39.8	530	6 Q9TSL8	Q9TSL8 macaca fasc
33	1113	39.8	529	6 Q18736	Q18736 bos taurus
34	1102.5	39.4	530	4 Q14928	Q14928 homo sapien
35	1101.5	39.4	530	4 Q9HAW7	Q9HAW7 homo sapien
36	1099.5	39.3	530	4 Q9HAW8	Q9HAW8 homo sapien
37	1092.5	39.0	531	6 Q9XS55	Q9XS55 ovis aries
38	1089.5	38.9	530	4 Q00473	Q00473 homo sapien
39	1088.5	38.9	530	11 Q8VD44	Q8VD44 ratcus norv
40	1083	38.7	530	4 Q00474	Q00474 homo sapien
41	1081.5	38.7	531	11 Q8VD43	Q8VD43 ratcus norv
42	1081.5	38.7	532	4 Q96TE7	Q96TE7 homo sapien
43	1077	38.5	532	6 Q9XS56	Q9XS56 ovis aries
44	1076.5	38.5	532	4 Q8WU04	Q8WU04 homo sapien
45	1075	38.4	498	11 Q8VCQ9	Q8VCQ9 mus musculu

ALIGNMENTS

RESULT 1	ID	Q8WN97	PRELIMINARY;	PRT;	528 AA.
AC	Q8WN97	Q8WN97			
DT	01-MAR-2002	(TRENBLREL. 20, Created)			
DT	01-MAR-2002	(TRENBLREL. 20, Last sequence update)			
DT	01-MAR-2003	(TRENBLREL. 23, Last annotation update)			
DE	UDP-glucuronosyltransferase 2B30.				
GN	UGT2B30.				
OS	Macaca fascicularis (Cray eating macaque) (Cynomolpus monkey).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;				
OC	Cercopitheidae; Macaca.				
OX	NCBI_TaxID=9541;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Girard C., Barbier O., Belanger A.;				
RT	"Structure of the monkey UDP-glucuronosyltransferase UGT2B30 gene and				
RT	characterization of the expressed enzyme which conjugates androgens,				
RT	estrogens and progestins, in addition to cortisol and aldosterone				
RT	metabolites."				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF401557; AAL60145.1; -				
DR	InterPro; IPR002213; UDP_gluco_trans.				
DR	Pfam; PF00201; UDPGT; 1.				
DR	PROSITE; PS00375; UDPGT; 1.				
KW	Transferase.				
SO	SEQUENCE 528 AA; 60366 MW; 361349D4717C68D5 CRC64;				
QY	Query Match	84.6%;	Score 2368.5;	DB 6;	Length 528;
QY	Best Local Similarity	83.2%;	Pred. No. 3.1e-180;		
QY	Matches 440;	Conservative 40;	Mismatches 48;	Indels 1;	Gaps 1;
DB	1 MSKMTSALLILQSCYFSSGCGKVLWPTFSHMWNKTIIDELVORGHVTVLASSA 60				
QY	1 MSKMTSALLILQSCYFSSGCGKVLWPTFSHMWNKTIIDELVORGHVTVLASSA 60				
QY	61 SIFDPNPSSTLKEFVYVSLTKTEFEDIKOLVKNWAEPLPKDTFMSYFQVOEIMWTFN 120				

```

Db 61 SILPGNNPNSALKFEICPTSLTETEFEDSVTLQVKNMSDIPKDETFPHFLQVQEMMTYTG 120
Qy 121 DILKFCDDIYVSNKKLMKKQLQESRFDVVLADAVPFGGLAELKIPVYSLRSPGYAI 180
Db 121 DMIRFCDDVSNKKLMKKQLQESRFDVVLADAIISGGLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Db 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Qy 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFGGLHCKPAKLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFGGLHCKPAKLPKMEEFVQSSG 300
Qy 301 ENGUVVSLGSMVNTSEERANVTASALAKIPQKVLMPDQNKPTGLNTRLKYMPION 360
Db 301 ENGUVVSLGSMVNTSEERANVTASALAKIPQKVLMPDQNKPTGLNTRLKYMPION 360
Qy 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKTM 420
Db 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKTM 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSPAA 480
Db 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSPAA 480
Qy 481 HDLTFQHSIDVIGFLLTCVATAIFLTCKCFSPCKQFNKTRKIEKRE 529
Db 481 HDLTFQHSIDVIGFLLTCVATAIFLTCKCFSPCKQFNKTRKIEKRE 529

```

RESULT 2

```

ID 097951 PRELIMINARY; PRT: 529 AA.
AC 097951;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase.
GN UGT2B18.
OS Macaca fascicularis (Crab eating macaque) (Synonym: J. monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN (1)
RP SEQUENCE FROM N.A.
RA Beaulieu M., Levesque E., Barbier O., Turgeon D., Belanger G.,
RA Hum D.W., Belanger A.;
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL, AF016310, AAC96726.1, -.
DR InterPro: IPR002213, UDP_gluco_trans.
DR Pfam: PF00201, UDPGT, 1.
DR PROSITE: PS00375, UDPGT, 1.
DR Transferase.
SQ SEQUENCE 529 AA, 60801 MW, 3CECB49786C3601F CRC64;

```

Query Match 82.6%; Score 2311; DB 6; Length 529;
 Best Local Similarity 81.3%; Pred. No. 1.2e-175;
 Matches 430; Conservative 39; Mismatches 60; Indels 0; Gaps 0;

```

Qy 1 MSMKNTSALLIQLSCYSSGCGKVLVMPTEFSHMNIKITLDELVORGEHVTLASSA 60
Db 1 MSMKNTSALLIQLSCYSSGCGKVLVMAAESHMMNKTILBELVORGEHVTLASSA 60
Qy 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTFWSYFSQVQOELMTFN 120
Db 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTFWSYFSQVQOELMTFN 120
Qy 121 DILKFCDDIYVSNKKLMKKQLQESRFDVVLADAVPFGGLAELKIPVYSLRSPGYAI 180
Db 121 DILKFCDDIYVSNKKLMKKQLQESRFDVVLADAVPFGGLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Db 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240

```

```

Qy 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Db 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Qy 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFGGLHCKPAKLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFGGLHCKPAKLPKMEEFVQSSG 300
Qy 301 ENGUVVSLGSMVNTSEERANVTASALAKIPQKVLMPDQNKPTGLNTRLKYMPION 360
Db 301 ENGUVVSLGSMVNTSEERANVTASALAKIPQKVLMPDQNKPTGLNTRLKYMPION 360
Qy 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKTM 420
Db 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKTM 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSPAA 480
Db 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSPAA 480
Qy 481 HDLTFQHSIDVIGFLLTCVATAIFLTCKCFSPCKQFNKTRKIEKRE 529
Db 481 HDLTFQHSIDVIGFLLTCVATAIFLTCKCFSPCKQFNKTRKIEKRE 529

```

RESULT 3

```

ID 09GLD9 PRELIMINARY; PRT: 529 AA.
AC 09GLD9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase UGT2B33.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN (1)
RP SEQUENCE FROM N.A.
RA TISUB=Liver;
RA Dean B., Zhao S., King C.;
RT "Nucleotide and Deduced Amino Acid Sequence of UGT2B33: a Novel UDP-
RT Glucuronosyltransferase Isoform Cloned from Female Rhesus Monkey
RT Liver ";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL, AF294902, AAG21378.1, -.
DR InterPro: IPR002213, UDP_gluco_trans.
DR Pfam: PF00201, UDPGT, 1.
DR PROSITE: PS00375, UDPGT, 1.
DR Transferase.
SQ SEQUENCE 529 AA, 60858 MW, B1956F947F1F78D6 CRC64;

```

Query Match 81.9%; Score 2291; DB 6; Length 529;
 Best Local Similarity 81.0%; Pred. No. 4.7e-174;
 Matches 427; Conservative 41; Mismatches 57; Indels 2; Gaps 1;

```

Qy 1 MSMKNTSALLIQLSCYSSGCGKVLVMPTEFSHMNIKITLDELVORGEHVTLASSA 60
Db 1 MSMKNTSALLIQLSCYSSGCGKVLVMAAESHMMNKTILBELVORGEHVTLASSA 60
Qy 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTFWSYFSQVQOELMTFN 120
Db 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTFWSYFSQVQOELMTFN 120
Qy 121 DILKFCDDIYVSNKKLMKKQLQESRFDVVLADAVPFGGLAELKIPVYSLRSPGYAI 180
Db 121 DILKFCDDIYVSNKKLMKKQLQESRFDVVLADAVPFGGLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240
Db 181 EKHSGLLFPSPSYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQFYSVL 240

```

```
Qy 241 GRPTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRHTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
Qy 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
Qy 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDNLNIAHKAAGAVERINFETM 420
Db 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDNLNIAHKAAGAVERINFETM 420
Qy 421 TSEDLALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Db 421 SSTDANLAKTVINDPLYKENMKLSRIQDQPVKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Qy 481 HDLTFPQHSIDVIGFLITCVATAIFLTKCFELSCQKFNKTRIEK 527
Db 481 HDLTFPQHSIDVIGFLITCVATAIFLTKCFELSCQKFNKTRIEK 527
```

```
RESULT 4
Q9GLE0 PRELIMINARY; PRT; 529 AA.
AC Q9GLE0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase UGT2B9*2.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
NCBI_TaxID=9544;
RX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Dean B., Zhao S., King C.;
RT "Nucleotide and Deduced Amino Acid Sequence of UGT2B9*2: a Novel UDP-
RT Glucuronosyltransferase Isoform Cloned from Female Rhesus Monkey
RT Liver."
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF294901; AAC21377.1; -.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
DR KMW
DR TRANSFERASE.
SQ SEQUENCE 529 AA; 60831 MW; 1225A0163D9AADF2 CRC64;
```

Query Match 81.8%; Score 2289; DB 6; Length 529;
Best Local Similarity 80.5%; Pred. No. 6.8e-174; Indels 0; Gaps 0;
Matches 426; Conservative 38; Mismatches 65;

```
Qy 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFESHMNMKTIIDELVORGEVTVLASSA 60
Db 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFESHMNMKTIIDELVORGEVTVLASSA 60
Qy 61 SISDPNSPSTLKEVEYVVSITKTEFEDIKQLVKRWAEIPKDTFWSYFSQVQBITMTFN 120
Db 61 SISDPNSPSTLKEVEYVVSITKTEFEDIKQLVKRWAEIPKDTFWSYFSQVQBITMTFN 120
Qy 121 DILKPKCDIYVSNKKLKKLOESRFDVADVPFPGELAEELKIPFVSLRSPGYAI 180
Db 121 DILKPKCDIYVSNKKLKKLOESRFDVADVPFPGELAEELKIPFVSLRSPGYAI 180
Qy 181 EKHSGGLFPSPSYVPMVMSLSDQMTFIERVKNMIVYLVEFVFOIPFMKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVPMVMSLSDQMTFIERVKNMIVYLVEFVFOIPFMKMDQFYSVL 240
Qy 241 GRPTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRPTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
```

```
Qy 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
Qy 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDNLNIAHKAAGAVERINFETM 420
Db 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDNLNIAHKAAGAVERINFETM 420
Qy 421 TSEDLALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Db 421 SSTDANLAKTVINDPLYKENMKLSRIQDQPVKPLDRAVFWIEFVNRHKGAKHLRSAA 480
Qy 481 HDLTFPQHSIDVIGFLITCVATAIFLTKCFELSCQKFNKTRIEK 529
Db 481 HDLTFPQHSIDVIGFLITCVATAIFLTKCFELSCQKFNKTRIEK 529
```

```
RESULT 5
Q9TSL6 PRELIMINARY; PRT; 529 AA.
ID Q9TSL6
AC Q9TSL6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase 2B23 (EC 2.4.1.17).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20043918; PubMed=10579317;
RA Barbier O., Levesque E., Belanger A., Hum D.W.;
RT "UGT2B23, a novel uridine diphosphate-glucuronosyltransferase enzyme
RT expressed in steroid target tissues that conjugates androgen and
RT estrogen metabolites."
RL Endocrinology 140:5538-5548 (1999).
DR EMBL; AF112113; AAF14353.1; -.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
DR KMW
DR GLYCOSYLTRANSFERASE; TRANSFERASE.
SQ SEQUENCE 529 AA; 60957 MW; 646315B6D970541A CRC64;
```

Query Match 80.8%; Score 2261; DB 6; Length 529;
Best Local Similarity 79.2%; Pred. No. 1.2e-171; Indels 0; Gaps 0;
Matches 419; Conservative 42; Mismatches 68;

```
Qy 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFESHMNMKTIIDELVORGEVTVLASSA 60
Db 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFESHMNMKTIIDELVORGEVTVLASSA 60
Qy 61 SISDPNSPSTLKEVEYVVSITKTEFEDIKQLVKRWAEIPKDTFWSYFSQVQBITMTFN 120
Db 61 SISDPNSPSTLKEVEYVVSITKTEFEDIKQLVKRWAEIPKDTFWSYFSQVQBITMTFN 120
Qy 121 DILKPKCDIYVSNKKLKKLOESRFDVADVPFPGELAEELKIPFVSLRSPGYAI 180
Db 121 DILKPKCDIYVSNKKLKKLOESRFDVADVPFPGELAEELKIPFVSLRSPGYAI 180
Qy 181 EKHSGGLFPSPSYVPMVMSLSDQMTFIERVKNMIVYLVEFVFOIPFMKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVPMVMSLSDQMTFIERVKNMIVYLVEFVFOIPFMKMDQFYSVL 240
Qy 241 GRPTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
Db 241 GRHTLSTETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
Qy 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLMPFDGNKPTDGLNTRLYKWIPON 360
```

QY 361 DLGHPKTKAFITGGMNGIYEAIYHGVPMVGVPIDFGDQDQDNIAMKAKGAABEINFKT 420
DB 361 DLGHPKTKAFITGGMNGIYEAIYHGVPMVGIPLEFADQPDNIAMKTRGAALVDFTDM 420
QY 421 TSEDLRALRTVITDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKILBSAA 480
DB 421 SSTDLVNLKTVINDSPYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKILBSAA 480
QY 481 HDLTFQHSYIDVIGFLITCVATAIFLFTKCFPLFSCQKFNKTRKIERE 529
DB 481 HDLTFQHSYIDVIGFLITCVATAIFLFTKCFPLFSCQKFNKTRKIERE 529

RESULT 6

Q8R084 ID Q8R084 PRELIMINARY; PRT; 529 AA.
AC Q8R084;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP glycosyltransferase 2 family, polypeptide B17
GN (UDP-glucuronosyltransferase 2B1 precursor).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RL EMBL; BC027200; AAH27200.1; -
DR EMBL; AK050435; BAC34254.1; -
DR MGD; MGI:1919023; 1300012D2ORik.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 529 AA; 60157 MW; B13D713102F794DF CRC64;

Query Match 71.5%; Score 1999.5; DB 11; Length 529;
Best Local Similarity 69.9%; Pred. No. 8.1e-151;
Matches 369; Conservative 62; Mismatches 96; Indels 1; Gaps 1;
QY 1 MSMKMTSA-LLLIQLSCYFSSGCGKVLVWPTFESHMNIKITLDELVORGHVTVLASA 60
DB 1 MSMKQASVFLIIQFICIRPGACGKVLVWPTFESHMNIKITLDELVORGHVTVLASA 60
QY 61 SISDPNPSSTLKEFVYPSLTKTFEEDIIKOLYKRW-ELPKDTFWSYFSQVQEIIMTF 119
DB 61 SILGPSNNESSINFEIYSAPLSKDLEVAFEKVMGNMTELEKLPFWTSYSKLOKISSEY 120
QY 120 NDILRFKPCDVIYNKKMLKQESRPDVVLADAVPFGELLAELIKIPVYSLRSPGYA 179
DB 121 SDMKESFKAIVVNSGLMKKQSGKFDVVLADALVPCGELLSELKTVLVSLRCPGYK 180
QY 180 IEKSHSGGLFPSPSYVPMVMSLSDQMTFIERVKMNIYVLYFEFWQIFDMKKMQDFYS 239
DB 181 CEKYSGGGLPLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYFEFWQIFDMKKMQDFYS 240
QY 240 LGRPTTSETMAKADIWILIRYWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQS 299
DB 241 LGRPTTSETMAKADIWILIRYWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQS 300

QY 300 GENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 359
DB 301 GEHGVVVSLSGSMVKIKKEERANYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMIPO 360
QY 360 NDILGHPKTKAFITGGMNGIYEAIYHGVPMVGVPIDFGDQDQDNIAMKAKGAABEINFKT 419
DB 361 NDILGHPKTKAFITGGMNGIYEAIYHGVPMVGVPIDFGDQDQDNIAMKAKGAABEINFKT 420
QY 420 MTSDDLRLARTVITDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKILBSA 479
DB 421 MSTDDLRLARTVITDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKILBSA 480
QY 480 AHDLTWFQHSYIDVIGFLITCVATAIFLFTKCFPLFSCQKFNKTRKIERE 527
DB 481 AHDLTWFQHSYIDVIGFLITCVATAIFLFTKCFPLFSCQKFNKTRKIERE 527

RESULT 7

Q8K154 ID Q8K154 PRELIMINARY; PRT; 532 AA.
AC Q8K154;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP glycosyltransferase 2 family, polypeptide B4.
GN A1788959.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RX EMBL; BC028826; AAH28826.1; -
DR MGD; MGI:2140962; A1788959.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 532 AA; 60877 MW; 82967F4EFF910A0B CRC64;

Query Match 71.2%; Score 1993.5; DB 11; Length 532;
Best Local Similarity 68.6%; Pred. No. 2.5e-150;
Matches 365; Conservative 78; Mismatches 86; Indels 3; Gaps 2;
QY 1 MSMKMTSA-LLLIQLSCYFSSGCGKVLVWPTFESHMNIKITLDELVORGHVTVLASA 58
DB 1 MPVGMTAALLLIQLSGFSGTGKVLVWPMFESHMNIKITLDELVORGHVTVLASA 60
QY 59 SASISFPNPSSTLKEFVYPSLTKTFEEDIIKOLYKRW-ELPKDTFWSYFSQVQEIIMTF 117
DB 61 SASISYEVDNLSAIEFTYPSYLSLSELEIFWLSLKYIYELPKOSFWGFWLQEMWV 120
QY 118 TENDILKPCDVIYNKKMLKQESRPDVVLADAVPFGELLAELIKIPVYSLRSPG 177
DB 121 VDSKTFESLCKDVFNKELMTKQKSRFDVILADPFCGDLLEVLKIPVYSLRFPG 180
QY 178 YAEKSHSGGLFPSPSYVPMVMSLSDQMTFIERVKMNIYVLYFEFWQIFDMKKMQDFYS 237
DB 181 STYKYSGGGLPLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYFEFWQIFDMKKMQDFYS 240
QY 238 EVLGRPTTSETMAKADIWILIRYWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQ 297
DB 241 EVLGRPTTSETMAKADIWILIRYWDPOFPHPLIPNVEFVGLHCKPAKPLPKMEEFVQ 300
QY 298 SSGENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMI 357
DB 301 SSGENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGNKPDTLGLNTRLYKMI 360
QY 358 PNDLGHKPKTKAFITGGMNGIYEAIYHGVPMVGVPIDFGDQDQDNIAMKAKGAABEINF 417

Db 361 PONDLLGHSKTRAFITTHGNGIYEAIYHGIPVGIPLFGQYDNIVLKKAKGAVRUDF 420
Qy 418 KTMSEDLRALRTVITDSSYKENAMRLSRTHHOQPVKPLDRAVFWIEFWNRHKGAKHLR 477
Db 421 LTMSTDLHTLTKVTVNDPSYKENAMRLSRTHHOQPVKPLDRAVFWIEFWNRHKGAKHLR 480
Qy 478 SAADLTWFQHSYIDVIGFLITCVATAIFLFTKCPFLSCQKFNKTRKIEKRE 529
Db 481 VAARDLSWQYHSLDVLGFLACVLTWFIILKCKLFCQKLTGAKRKKE 532

RESULT 8

Q8VIF9 PRELIMINARY; PRT; 528 AA.
AC O8VIF9:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase 2821 (EC 2.7.1.17).
GN UGT2B21.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N. A.
RC STPAIN=Hartley; TISSUE=Liver;
RA Ishii Y., Miyoshi A., Tsuda K., Tsuda M., Nagamatsu Y., Yoshisue K.,
Tanaka M., Matsuno S., Ohgita S., Oguri K.;
RT "Hetero-oligomer formation of UDP-glucuronosyltransferases determines
its substrate specificity: Metabolic activation of morphine to
morphine-6-glucuronide by UGT2B21 and UGT2B22 hetero-oligomer.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB034987; BAB82476.1; -
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 528 AA; 60945 MW; 534D7BAABA67ACB CRC64;

Query Match 70.4%; Score 1970.5; DB 11; Length 528;
Best Local Similarity 68.4%; Pred. No. 1.7e-148;
Matches 361; Conservative 73; Mismatches 92; Indels 1; Gaps 1;
Qy 3 MKMTSALLILLOSCYFSSGCGKVLWPTESHMNITITIDELVORGEVTVLASSA 62
Db 1 MKRIALLLLQLCHFHSGCGKVLWPMESHMNITIDELVIRGHEVTVLRPSCFI 60
Qy 63 SFDNSPSTLKEEYVPSVLTKEFEDIKOLVKRAELPK-DTFMSYFSQVOEIMWTF 121
Db 61 FVDVNTTSEIFETFFHNSFTRDYWEKITDVTWTLNGSVDTCLDPPEVEKLFKHSF 120
Qy 122 ILRKFCKDIVSNKMLKQLQESRPDVLAADVPPGELLALBLKIPVYSIRFSPGAYE 181
Db 121 EMEVNCXELVSNKKFMKVLQESRFDILADAVPGELVAELIHPVYSIRFSGPQAE 180
Qy 182 KHSGGLPPSPYVYVWSELSDQMTFIRVKNMIVLVYFEFVQIFDMKKNDQYSEVLG 241
Db 181 KRAGGLPPSPYVYVWSELSDQMTFIRVKNMIVLVYFEFVQIFDMKKNDQYSEVLG 240
Qy 242 RPTLTSETMAKADILINRYWDFOPPHPLPVEVFGHCKPAKPLPKEMEFPVQSSGE 301
Db 241 KPSTLLETMSKADMLINSYDMEPHPSLPNFDITGLHCKPAKPLPKEMEFPVQSSGE 300
Qy 302 NGVVVFSIGSVNSNTSEERANVIASALAKIPQKVLMPFDGKPDITLGINTRLYKWIPO 361
Db 301 HGIVVFSIGSVNMTDEKANLIALAQIPQKVLMPFDGKPDITLGINTRLYKWIPO 360
Qy 362 LLGHPKTAFTTHGANGIYEAIYHGVWGVPIFGDQDNIAMHKAAGAIVEINFKTMT 421
Db 361 LLGHPKTAFTTHGANGIYEAIYHGVWGVPIFGDQDNIAMHKAAGAIVEINFKTMT 420

Qy 422 SEDLLRALRTVITDSSYKENAMRLSRTHHOQPVKPLDRAVFWIEFWNRHKGAKHLRSAH 481
Db 421 STDLLNKLKTVINNPYSYKENAMRLSRTHHOQPVKPLDRAVFWIEFWNRHKGAKHLRLAH 480
Qy 482 DLTWFQHSYIDVIGFLITCVATAIFLFTKCPFLSCQKFNKTRKIEKRE 529
Db 481 NLTWQYHSLDVLGFLACVATAIFLFTKCPFLSCQKFNKTRKIEKRE 528

RESULT 9

Q8K169 PRELIMINARY; PRT; 530 AA.
AC Q8K169:
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase 2 family, member 5.
GN UGT2B5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC028262; AAH28262.1; -
DR MGI; MGI:98900; UGT2B5.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 60874 MW; 9FEFAA9ABE861D9 CRC64;

Query Match 68.2%; Score 1908.5; DB 11; Length 530;
Best Local Similarity 67.0%; Pred. No. 1.5e-143;
Matches 355; Conservative 66; Mismatches 108; Indels 1; Gaps 1;
Qy 1 MSMTSALLILLOSCYFSSGCGKVLWPTESHMNITITIDELVORGEVTVLASSA 60
Db 1 MPGMISALLLLQISCFRSVCGKVLWPMESHMNITITIDELVORGEVTVLRPSA 60
Qy 61 SISDPSPSTLKEEYVPSVLTKEFEDIKOLVKRA-ELPKDTFMSYFSQVOEIMWTF 119
Db 61 YVUDPPKSGGLKRETFPTVSQKDLNENFKFVDVNTTSEIFETFFHNSFTRDYWEKITDVTWTLNGSVDTCLDPPEVEKLFKHSF 120
Qy 120 NDILRKFCKDIVSNKMLKQLQESRPDVLAADVPPGELLALBLKIPVYSIRFSPGAYE 179
Db 121 SDVPLSLCKDVSNKMLKQLQESKFDVLLSDPVASGELLALBLQIPVYSIRFSPGYQ 180
Qy 180 IEXHSGGLPPSPYVYVWSELSDQMTFIRVKNMIVLVYFEFVQIFDMKKNDQYSEVLG 239
Db 181 IEXSGGFLPPSPYVYVWSELSDQMTFIRVKNMIVLVYFEFVQIFDMKKNDQYSEVLG 240
Qy 240 LGRPITLSETMAKADILINRYWDFOPPHPLPVEVFGHCKPAKPLPKEMEFPVQSS 299
Db 241 LGRPTLLETMAKADILINRYWDFOPPHPLPVEVFGHCKPAKPLPKEMEFPVQSS 300
Qy 300 GENGVVFSIGSVNSNTSEERANVIASALAKIPQKVLMPFDGKPDITLGINTRLYKWIPO 359
Db 301 GDHGVVFSIGSVNSNTSEERANVIASALAKIPQKVLMPFDGKPDITLGINTRLYKWIPO 360
Qy 360 NDILGHPKTAFTTHGANGIYEAIYHGVWGVPIFGDQDNIAMHKAAGAIVEINFKTMT 419
Db 361 NDILGHPKTAFTTHGANGIYEAIYHGVWGVPIFGDQDNIAMHKAAGAIVEINFKTMT 420
Qy 420 MTSDDLRLALRTVITDSSYKENAMRLSRTHHOQPVKPLDRAVFWIEFWNRHKGAKHLRSA 479
Db 421 MTSKDVNLALBEYIENFPYKNAAMLSITHDQPKPLDRAVFWIEFWNRHKGAKHLRPL 480
Qy 480 AHDLTWFQHSYIDVIGFLITCVATAIFLFTKCPFLSCQKFNKTRKIEKRE 529

Db 481 GHNLTWQYHSLDVIIGFLSCVATTIVLSCVKKLLFYRFVYKKNKMKNE 530

RESULT 10

Q91WH2 PRELIMINARY; PRT; 530 AA.

ID Q91WH2 AC Q91WH2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Similar to UDP-glucuronosyltransferase 2 family, member 5.
GN AA986709.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Kidney;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; BC015272; AAH15272.1;
DR EMBL; AK034801; BAC28835.1;
DR MGI; MGI:2140794; AA986709.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
KW Transferrase.
SQ SEQUENCE 530 AA; 60929 MW; B475642270B28EF1 CRC64;

Query Match 67.7%; Score 1895.5; DB 11; Length 530;
Best Local Similarity 66.8%; Pred. No. 1.6e-142;
Matches 354; Conservative 64; Mismatches 111; Indels 1; Gaps 1;

Qy 1 MSMKWTALLLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTVLASSA 60
Db 1 MPKGISALLLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTVLASSA 60

Qy 61 SISFDPNPSTLKEFVYVSVLTKTEFEDIIKQLVKRA-ELPKDTFWSYFSQVGEIMWTF 119
Db 61 YVLDPKKSPGLKFTFTSVSKDNLENFFIKFVDMVTYEMPRDYLCLSYSLQNMIDF 120

Qy 120 NDILRPFCKDIYSNKKMLKQESRPDVVLADAVPFGELLALIKIPVYSLRSPGYA 179
Db 121 SDYLSLCKEAVSNKQMTKQESKFDVLLSDPVASCCELLAEIQLPFLYSIRSPGYQ 180

Qy 180 IEKHSGLLPSPSYVPMVSELSDQMTFIERVKMNIYLYFEFQIPLFDMKKMDFSEV 239
Db 181 IEKSGRGLLPSPSYVPMVSELSDQMTFIERVKMNIYLYFEFQIPLFDMKKMDFSEV 240

Qy 240 LGRPTTLSETMAKADIWLIRYWDFOFPHPLPNVEFVGGHCKCPAKPLPKMEEFVQSS 299
Db 241 LGRPTTLSETMAKADIWLIRYWDFOFPHPLPNVEFVGGHCKCPAKPLPKMEEFVQSS 300

Qy 300 GENGIVVSLGSMVNTSEERANVIALAKIPQKVLWRFDPGNKDTGLNTRLYKWIPO 359
Db 301 GDHGVVSLGSMVNTSEERANVIALAKIPQKVLWRFDPGNKDTGLNTRLYKWIPO 360

Qy 360 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFT 419
Db 361 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFT 420

Qy 420 MTSDDLRLALRTVTITDSYKENARLSRIHHDQVVKPLDRAVFWIEFVRRHKGAKHLSA 479
Db 421 MTSDDLRLALRTVTITDSYKENARLSRIHHDQVVKPLDRAVFWIEFVRRHKGAKHLSA 480

Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFELSCCKENKFKIEKRE 529
Db 481 GHNLTWQYHSLDVIIGFLSCVATTIVLSCVKKLLFYRFVYKKNKMKNE 530

RESULT 11

Q8BLI9 PRELIMINARY; PRT; 529 AA.

ID Q8BLI9 AC Q8BLI9;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Similar to UDP-glucuronosyltransferase 2B5 precursor.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK083294; BAC28847.1;
SQ SEQUENCE 529 AA; 60453 MW; 7C246E0716957469 CRC64;

Query Match 67.7%; Score 1895; DB 11; Length 529;
Best Local Similarity 66.4%; Pred. No. 1.7e-142;
Matches 352; Conservative 65; Mismatches 111; Indels 2; Gaps 2;

Qy 1 MSMKWTALLLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTVLASSA 60
Db 1 MPKGISALLLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTVLASSA 60

Qy 61 SISFDPNPSTLKEFVYVSVLTKTEFEDIIKQLVKRA-ELPKDTFWSYFSQVGEIMWTF 119
Db 61 YVLDPKKSPGLKFTFTSVSKDNLENFFIKFVDMVTYEMPRDYLCLSYSLQNMIDF 120

Qy 120 NDILRPFCKDIYSNKKMLKQESRPDVVLADAVPFGELLALIKIPVYSLRSPGYA 179
Db 121 SDYLSLCKEAVSNKQMTKQESKFDVLLSDPVASCCELLAEIQLPFLYSIRSPGYQ 179

Qy 180 IEKHSGLLPSPSYVPMVSELSDQMTFIERVKMNIYLYFEFQIPLFDMKKMDFSEV 239
Db 180 IEKSGRGLLPSPSYVPMVSELSDQMTFIERVKMNIYLYFEFQIPLFDMKKMDFSEV 239

Qy 240 LGRPTTLSETMAKADIWLIRYWDFOFPHPLPNVEFVGGHCKCPAKPLPKMEEFVQSS 299
Db 240 LGRPTTLSETMAKADIWLIRYWDFOFPHPLPNVEFVGGHCKCPAKPLPKMEEFVQSS 299

Qy 300 GENGIVVSLGSMVNTSEERANVIALAKIPQKVLWRFDPGNKDTGLNTRLYKWIPO 359
Db 300 GENGIVVSLGSMVNTSEERANVIALAKIPQKVLWRFDPGNKDTGLNTRLYKWIPO 359

Qy 360 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFT 419
Db 360 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFT 419

Qy 420 MTSDDLRLALRTVTITDSYKENARLSRIHHDQVVKPLDRAVFWIEFVRRHKGAKHLSA 479
Db 420 MTSDDLRLALRTVTITDSYKENARLSRIHHDQVVKPLDRAVFWIEFVRRHKGAKHLSA 479

Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFELSCCKENKFKIEKRE 529
Db 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFELSCCKENKFKIEKRE 529

RESULT 12

Q8VIF8

RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=Hartley; Tissue=Liver;
 RX MEDLINE=99410697; PubMed=10479484;
 RA Smith S.A., Nagalla S.R., Andrews D.P., Olsen G.D.;
 RT "Morphine regulation of a novel uridine diphosphate glucuronosyl-
 transferase in guinea pig pups following in utero exposure";
 RL Mol. Genet. Metab. 68:68-77(1999).
 DR EMBL: AF175221; AAD51732.1; -;
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam: PF00201; UDPGT; 1.
 DR PROSITE: PS00375; UDPGT; 1.
 KM Transferase.
 SQ SEQUENCE 530 AA; 59895 MW; EE2F394D3FD484E1 CRC64;

 Query Match 64.4%; Score 1802; DB 11; Length 530;
 Best Local Similarity 65.3%; Pred. No. 4.5e-135;
 Matches 344; Conservative 63; Mismatches 118; Indels 2; Gaps 2;

 QY 4 KMTSALLIOLSCYSSGCGKLVMPTEFSHMNIKITIDELVORGHVTVLASSASIS 63
 DB 5 KLASAVLLDLCC-AGSGFCGKLVPCGSMHMLKTLLELVKRGHEVTVLTSNNLF 63
 QY 64 FDPNSPSTLKEVYFVSLTKTEFEDIIKQLYKRAE-LPKDTFMSYFSQVQEIWMTFNDI 122
 DB 64 IDYNHRAFNFEVIVPTDKMSSENIINFEI LAVNMPWPLMQSGKLLQGFVQITED 123
 QY 123 LRKFKQIVSNKKMLKQOESRFDVVLADAVFPFGEELAEELIKFPVYSLRSPGIALEK 182
 DB 124 LGLNCRNTVYVQSLMKKLRDSKYDLVLPVPCGELVAEMLVGPVMMLKFSMGHTIEK 183
 QY 183 HSGGLLPSPSYVYVMSSELSDQMTFIERKNNIYVLYFEFVQIDPMKKWDQFYSEVGR 242
 DB 184 YCGQLRAPSPYVPLVGLGILTTMTMERKNNVSEVLDFWLOQDYKFWDFSEALGR 243
 QY 243 PTLSETMAKADILIRNYWDFQPHPLPNVEFVGLHCKPAKPLPKMEBEFVOSSEN 302
 DB 244 PTLCEINGKAIWIRTYWDFEPRPPLPNEFVGLHCKPAKPLPKMEBEFVOSSED 303
 QY 303 GVVVNSLSGMSNTSEERANVASALAKIPQKYLRFPGNKEDTGLNTRLYKMIPONDL 362
 DB 304 GVVVNSLSGMSVNTSEERANVASALAKIPQKYLRFPGNKEDTGLNTRLYKMIPONDL 363
 QY 363 LGHPRTKAFITHGNGNGIYEALYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMS 422
 DB 364 LGHPRTKAFITHGNGNGIYEALYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMS 423
 QY 423 EDLRLALRTVITDSSYKENAMRLSRIHHDQVYKPLDRAVFWIEFVNRHKGAGLRSAAH 482
 DB 424 ADLLGALRTVINDPYKENAMRLSRIHHDQVYKPLDRAVFWIEFVNRHKGAGLRSAAH 483
 QY 483 LTFPHYSIDVIGFLLTCVATAIPLFTKCFLESCOKFNKTRKIEKE 529
 DB 484 LSWFQYHSIDVIGFLLTCVATAIPLFTKCFLESCOKFNKTRKIEKE 530

 RESULT 15
 Q9Y4X1 PRELIMINARY; PRT; 527 AA.
 AC Q9Y4X1;
 DT 01-NOV-1999 (TRENBLREL. 12, Created)
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE UDP glucuronosyltransferase (EC 2.4.1.17).
 GN UGT2A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC Tissue=Olfactory organ;
 RX MEDLINE=99289328; PubMed=10359671;

RA Jedischny G.A., Cassidy A.J., Sales M., Pratt N., Burchell B.;
 RT "Cloning and characterization of a novel human olfactory UDP-
 glucuronosyltransferase";
 RL Biochem. J. 340:837-843(1999).
 DR EMBL: AJ006054; CAB41974.1; -;
 DR Genew; HGNC:12542; UGT2A1.
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam: PF00201; UDPGT; 1.
 DR PROSITE: PS00375; UDPGT; 1.
 KM Glucosyltransferase; Transferase.
 SQ SEQUENCE 527 AA; 59872 MW; B99E73AA9A421F17 CRC64;

 Query Match 64.2%; Score 1795.5; DB 4; Length 527;
 Best Local Similarity 65.0%; Pred. No. 1.5e-134;
 Matches 330; Conservative 72; Mismatches 103; Indels 3; Gaps 2;

 QY 24 GKVLWPTFESHMNIKITIDELVORGHVTVLASSASISFDPNSPSTLKEVYFVSLTK 83
 DB 21 GNVLIWPEGSHMNVKIIIDELIKKEHNVTVLAVASGALFITPSPNSLTFEIVKVPFGK 80
 QY 84 TEFEDIIKQLYKRAE-LPKDTFMSYFSQVQEIWMTFNDILRKFKQDIVSNKKMLKQ 141
 DB 81 ERIGVLIKDFVSTVLENRPSPSTIWRFYQEMAKVIKDFHVSQEI CDGVLKNQQLMALTK 140
 QY 142 ESREVDVLADAVFPFGEELAEELIKFPVYSLRSPGIALEKSGGLFPSPSYVPMVMSL 201
 DB 141 KSKFEVLSDVPFCGDI VALKGI PFMYSLRFPASTVEHGCKVYPRPVYAVVSEL 200
 QY 202 SDQMTFIERKNNIYVLYFEFVQIDPMKKWDQFYSEVLRPTLSETMAKADILIRNY 261
 DB 201 TDQMSFTDRIRNFISYHLODYMFETL-WKSDSYYSKALGRPTLCEFMGAEIWLIRTY 259
 QY 262 WDPQRPPLLPNVEFVGLHCKPAKPLPKMEBEFVOSSENGVVVFSLSGMSVNTSEERA 321
 DB 260 WDFEPRPPLPNVEFVGLHCKPAKPLPKMEBEFVOSSENGVVVFSLSGMSVNTSEKA 319
 QY 322 NVIASALAKIPQKYLRFPGNKEDTGLNTRLYKMIPONDLGHPTKAFITHGNGIY 381
 DB 320 NLIASALAKIPQKYLRFPGNKEDTGLNTRLYKMIPONDLGHPTKAFITHGNGIY 379
 QY 382 EAIYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMSSEDLRLALRTVITDSSYKEN 441
 DB 380 EAIYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMSSEDLRLALRTVITDSSYKEN 439
 QY 442 AMRLSRIHHDQVYKPLDRAVFWIEFVNRHKGAGLRSAAHDTWFOHYSIDVIGFLLVCV 501
 DB 440 AMRLSRIHHDQVYKPLDRAVFWIEFVNRHKGAGLRSAAHDTWFOHYSIDVIGFLLVCV 499
 QY 502 ATAIFLFTKCFLESCOKFNKTRKIEKE 529
 DB 500 TTAIFLVIGCCLFSCQKFGKIGKKKKE 527

 Search completed: December 5, 2003, 09:49:43
 Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:55 ; Search time 18 seconds
(without alignments)
1382.063 Million cell updates/sec

Title: US-09-980-729b-5
Perfect score: 2798
Sequence: 1 MSMKMTSALLILQLSCYFSS.....KCFLFSCQKFKTRIKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2602.5	93.0	528	1	UDB4_HUMAN
2	2369.5	84.7	528	1	UDB4_HUMAN
3	2330	83.3	529	1	UDB7_HUMAN
4	2297	82.1	529	1	UDB7_HUMAN
5	2292	81.9	529	1	UDBB_HUMAN
6	2281.5	81.5	529	1	UDBA_HUMAN
7	2253	80.5	529	1	UDBS_HUMAN
8	2169.5	77.5	530	1	UDBH_HUMAN
9	2158.5	77.1	530	1	UDBF_HUMAN
10	2100.5	75.1	530	1	UDBE_HUMAN
11	1991.5	71.2	530	1	UDBE_HUMAN
12	1973.5	70.5	529	1	UDBI_HUMAN
13	1938	69.3	531	1	UDBD_HUMAN
14	1938	69.3	531	1	UDBD_HUMAN
15	1930.5	69.0	530	1	UDBC_HUMAN
16	1904.5	68.1	530	1	UDBS_MOUSE
17	1894.5	67.7	530	1	UDB3_HUMAN
18	1856.5	66.4	530	1	UDB2_HUMAN
19	1840.5	65.8	530	1	UDB6_HUMAN
20	1777.5	63.0	530	1	UDB8_HUMAN
21	1761.5	63.0	527	1	UDB1_HUMAN
22	1593	56.9	502	1	UDB1_HUMAN
23	1201.5	42.9	531	1	UDB5_HUMAN
24	1196.5	42.8	533	1	UDB2_HUMAN
25	1188.5	41.5	533	1	UDB1_MOUSE
26	1158	41.4	535	1	UDB1_MOUSE
27	1137.5	40.7	531	1	UDB1_HUMAN
28	1137.5	40.7	533	1	UDB1_HUMAN
29	1135.5	40.6	535	1	UDB1_HUMAN
30	1128	40.3	534	1	UDB1_HUMAN
31	1126	40.2	534	1	UDB1_HUMAN
32	1124	40.2	534	1	UDB1_HUMAN
33	1120.5	40.0	530	1	UDB1_HUMAN

34	1120.5	40.0	530	1	UDB1_HUMAN	O6656 homo sapien
35	1109.5	39.7	532	1	UDB1_RABIT	O28612 oryctolagus
36	1099.5	39.3	530	1	UDB1_HUMAN	O9haw9 homo sapien
37	1098.5	39.3	530	1	UDB1_HUMAN	P36509 homo sapien
38	1094.5	39.1	530	1	UDB1_HUMAN	O9haw9 homo sapien
39	1094	38.7	520	1	UDB1_MOUSE	O62452 mus musculus
40	1081.5	38.5	531	1	UDB1_HUMAN	P19224 homo sapien
41	1078.5	38.5	531	1	UDB1_RAT	O64633 ratius norv
42	1077.5	38.5	530	1	UDB1_RAT	O64633 ratius norv
43	1066	38.1	531	1	UDB1_MOUSE	O64435 mus musculus
44	1045.5	37.4	529	1	UDB1_RAT	P08430 ratius norv
45	1045	37.3	531	1	UDB1_RABIT	O28611 oryctolagus

ALIGNMENTS

RESULT 1
UDB4_HUMAN STANDARD; PRT; 528 AA.
ID: UDB4_HUMAN
AC P06133; O60731; O60867; O75614; P36538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UDP-glucuronosyltransferase 2B4 precursor, microsomal (BC 2.4.1.17)
DE (UDP-GT) (Hydroxycholelic acid) (HLUG25) (UDP-GT-1).
GN UGT2B4 OR UGT2B11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87241362; PubMed=3109396;
RA Jackson M.R., McCarthy L.R., Harding D., Wilson S., Coughtrie M.W.H.,
RA Burchell B.;
RT "Cloning of a human liver microsomal UDP-glucuronosyltransferase
RT cDNA.";
RL Biochem. J. 242:581-586 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93326164; PubMed=8333863;
RA Jin C.-J., Miners J.O., Lillywhite K.J., McKenzie P.I.;
RT "cDNA cloning and expression of two new members of the human liver
RT UDP-glucuronosyltransferase 2B subfamily.";
RL Biochem. Biophys. Res. Commun. 194:496-503 (1993).
RN [3]
RP SEQUENCE FROM N.A., VARIANT GIU-458, AND CHARACTERIZATION.
RX MEDLINE=99303261; PubMed=10376768;
RA Levesque E., Beaulieu M., Hum D.W., Belanger A.;
RT "Characterization and substrate specificity of UGT2B4 (B458): a
RT UDP-glucuronosyltransferase encoded by a polymorphic gene.";
RL Pharmacogenetics 9:207-216 (1999).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANTS LEU-109 AND LEU-396.
RX McKenzie P.I.;
RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX Riedy M., Miller A.;
RT "Genomic organization and structure of the UGT2B gene complex at human
RT chromosome 4q13.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: UDP-GTs are of major importance in the conjugation and
CC subsequent elimination of potentially toxic xenobiotics and
CC endogenous compounds. This isozyme is active on polycyclic
CC estrogens (such as estril, 4-hydroxyestrone and 2-hydroxyestril)
CC and xenobiotics (such as 4-methylumbelliferone, 1-naphthol, 4-
CC nitrophenol, 2-aminophenol, 4-hydroxybiphenyl and menthol). It is
CC capable of 6 alpha-hydroxyglucuronidation of hydroxycholelic acid.
CC -i- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

OY		361	DLLGHPKTKAFITHGMNGVIGEALYYHGVPMWGPVIFPGPOLDNIIAHMKAKGAABVEINFKTM	420
OY		361	DLGHPRTRARITTCGANGLIYEALYYHGIPMWGVFLFADQPDNIIMHAKKAANVSLEDFRTM	420
OY		421	TSEDLBALRVRTTDSYSKENAMLSRIHHDPVKPLDRAVFWEIPEVRHKGAKHSAA	480
DB		421	SSTLDLNAIAKTIVINDPLYKENAMKLSRIHHDPVKPLDRAVFWEIPEVRHKGAKHLRVAA	480
OY		481	HDLTFPHOHSIDVIGFLTVCATXIFLFPTCKFLFSCOKFNKTRIEGRE	529
DB		481	HDLTWFPHSHSLDTVTGFLACVATYIFITTC-LFCWKRFVATGGKGRD	528
RESULT 2				
ID	UDBU_MACFA	STANDARD;	PRT; 528 AA.	
AC	OXYT55;			
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	UDP-glucuronosyltransferase 2B19 precursor, microsomal (EC 2.4.1.17).			
GN	UGT2B19.			
OS	Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).			
OC	Eukaryota; Metazoa; Chordata; Crniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;			
OC	Cercopithecinae; Macaca.			
OX	NCBI_TaxID=9541;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver, and Prostate:			
RX	MEDLINE=99203465; PubMed=10102998;			
RA	Belanger G., Barbier O., Hum D.W., Belanger A.;			
RT	"Molecular cloning, expression and characterization of a monkey			
RT	steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates			
RT	testosterone.";			
RU	Eur. J. Biochem. 260:701-708(1999).			
CC	-1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND			
CC	SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOGENICS AND			
CC	ENDOGENOUS COMPOUNDS. THIS ISOZYME DISPLAYS ACTIVITY TOWARD			
CC	SEVERAL CLASSES OF XENOBIOTIC SUBSTRATES: EUGENOL, 4-			
CC	METHYLBELLIFERONE, P-NITROPHENOL, 1-NAPTHOL, P,P'-BIPHENOL,			
CC	NARINEIN AND O,O'-BIPHENOL. ACTIVE ALSO ON 3A-HYDROXY AND 17B-			
CC	HYDROXY POSITIONS OF STEROIDS.			
CC	-1- FUNCTION: CONTRIBUTES TO THE FORMATION OF ANDROGEN GLUCURONIDE IN			
CC	EXTRAEPATIC STEROID TARGET TISSUES SUCH AS THE PROSTATE.			
CC	-1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor			
CC	beta-D-glucuronoside.			
CC	-1- SUBCELLULAR LOCATION: Microsomal (by similarity).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, OVARY, PROSTATE, COLON,			
CC	KIDNEY, PANCREAS, BRAIN, CEREBELLUM, MAMMARY GLAND AND EPIDIDYMIS.			
CC	NOT EXPRESSED IN SMALL INTESTINE, SPLEEN, BLADDER, ADRENAL GLAND			
CC	AND TESTIS.			
CC	-1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements			
CC	or send an email to license@ebi.ac.uk).			
CC	-----			
DR	EMBL; AF112112; AAC24435.1;			
DR	InterPro; IPR002213; UDP_Gluco_trans.			
DR	Pfam; PF00201; UDPGT_1.			
DR	PROSITE; PS00375; UDPGT_1.			
KW	Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;			
KW	MultiGene Family; Microsome.			
FT	SIGNAL	1..21	POTENTIAL.	
FT	CHAIN	22..528	UDP-GLUCURONOSYLTRANSFERASE 2B19.	
FT	TRANSMM	493..513	POTENTIAL (GLCNAC...) (POTENTIAL).	
FT	CARBOHYD	315..315	N-LINKED (GLCNAC...) (POTENTIAL).	
SO	SEQUENCE	528 AA;	3BFDA2E714A27AAE CC64;	

Query Match 84.7%; Score 2369.5; DB 1; Length 528;
 Best Local Similarity 82.8%; Pred. No. 7.5e-172;
 Matches 439; Conservative 43; Mismatches 47; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITIDELVORGHEVTLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITIDELVORGHEVTLAVST 60
 QY 61 SISFDPSPSTLAKFEVYVSLTTEPEDITIKOLVKRAELPKDTPFWSYFSQVOGIMMTFN 120
 DB 61 SILFDPNNSALKEICPTSLTEPEFQDSYVQLVKRMSDKRDTFMPFHLVQEMMTYTG 120
 QY 121 DILRKFKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 DB 121 DMIRKFKCDIVSNKLMKLOESRFDVVLADAIAPCEGLAELLIKIPVYSLRSPGYAL 180
 QY 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 DB 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 QY 241 GRPTTSETMAKADIVLIRNWDQFPHPLPNVEFGGLHCKRAKPLPKMEEFVYSSG 300
 DB 241 GRPTTLEIMAKAEMILIRNWDQFPHPLPNVEFGGLHCKRAKPLPKMEEFVYSSG 300
 QY 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 DB 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 QY 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDOLDNIAHMKAKGAAVEINFTKM 420
 DB 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDOLDNIAHMKAKGAAVEINFTKM 420
 QY 421 TSEDLALTRVITDSYKENAMRLSRHHDQPKYPLDRAVFVLEFVNRHKGAGHLASAA 480
 DB 421 SSTDLLALKVINDPIYKENAMKLSIHDDQPKYPLDRAVFVLEFVNRHKGAGHLASAA 480
 QY 481 HDLWFOHYSIDVIGPILTCVATAIPLFTKCFILFSCCKENRTIEKRE 529
 DB 481 HDLWFOHYSIDVIGPILTCVATAIPLFTKCFILFSCCKENRTIEKRE 529
 DB 481 HDLWFOHYSIDVIGPILTCVATAIPLFTKCFILFSCCKENRTIEKRE 529

RESULT 3
 UDB7 HUMAN STANDARD; PRT; 529 AA.

AC P16662;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glucuronosyltransferase 2B7 precursor, microsomal (EC 2.4.1.17)
 DE (UDPGT) (3,4-catechol estrogen specific) (UDPGT-2).
 GN UGT2B7 OR UGT1B2B9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90243659; PubMed=2159463;
 RA Raliter J.K., Sheen Y.Y., Owens I.S.;
 RT "Cloning and expression of human liver UDP-glucuronosyltransferase in
 RT COS-1 cells. 3,4-catechol estrogens and estradiol as primary
 RT substrates";
 RL J. Biol. Chem. 265:7900-7906(1990).
 RN [2]
 RP VARIANT UGT2B7*2.
 RX MEDLINE=21033487; PubMed=11186130;
 RA Bhasker C.R., McKinnon W., Stone A., Lo A.C., Kubota T., Ishizaki T.,
 RA Miners J.O.;
 RT "Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at
 RT amino acid 268: ethnic diversity of alleles and potential clinical
 RT significance.";

RL Pharmacogenetics 10:679-685(2000).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS.
 CC -1- FUNCTION: ITS UNIQUE SPECIFICITY FOR 3,4-CATECHOL ESTROGENS AND
 CC ESTRIOL SUGGESTS IT MAY PLAY AN IMPORTANT ROLE IN REGULATING THE
 CC LEVEL AND ACTIVITY OF THESE POTENT AND ACTIVE ESTROGEN
 CC METABOLITES.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC -1- beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, J05428; AAA36793.1; -;
 DR PIR, A35366; A35366.
 DR Genew; HGNC:12554; UGT2B7.
 DR MIM; 600068; -;
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0006629; P:lipid metabolism; TAS.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KM Multigene family; Microsome; Polymorphism.
 FT SIGNAL 1 23
 FT CHAIN 24 529
 FT TRANSEM 493 509
 FT CARBOHYD 67 67
 FT CARBOHYD 68 68
 FT CARBOHYD 315 315
 FT VARIANT 268 268
 FT H -> Y (IN ALLELE UGT2B7*2).
 FT FTID=VAR_012342.
 SQ SEQUENCE 529 AA; 60694 MW; 94B1CA6CE92P1446 CRC64;

Query Match 81.1%; Score 2330; DB 1; Length 529;
 Best Local Similarity 81.1%; Pred. No. 7.4e-169;
 Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITIDELVORGHEVTLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITIDELVORGHEVTLASSA 60
 QY 61 SISFDPSPSTLAKFEVYVSLTTEPEDITIKOLVKRAELPKDTPFWSYFSQVOGIMMTFN 120
 DB 61 SILFDPNNSALKEICPTSLTEPEFQDSYVQLVKRMSDKRDTFMPFHLVQEMMTYTG 120
 QY 121 DILRKFKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 DB 121 DMIRKFKCDIVSNKLMKLOESRFDVVLADAIAPCEGLAELLIKIPVYSLRSPGYAL 180
 QY 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 DB 181 EKHSGLLPSPSYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 QY 241 GRPTTSETMAKADIVLIRNWDQFPHPLPNVEFGGLHCKRAKPLPKMEEFVYSSG 300
 DB 241 GRPTTLEIMAKAEMILIRNWDQFPHPLPNVEFGGLHCKRAKPLPKMEEFVYSSG 300
 QY 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 DB 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 QY 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDOLDNIAHMKAKGAAVEINFTKM 420
 DB 361 DLGHPKTKAFITHGNGNGIYEALYHGVPMVGVPIFGDOLDNIAHMKAKGAAVEINFTKM 420

```

OY 61 S1SDPNSPLTKREYVPSVLTKKEPFDIIKOLVKRAAEI.PKDTFMSYFSQVOSIMMTFN 120
DB 61 S1LDFPNNSSLLKLEVEPDKITKTEFENISMOEYKRIETLPKDTFLYFSQMOZIMRRFG 120
OY 121 D1LRPFCKDVIYNSKKMLKKLOESRFVDVLDAVEPFCELLAEELKIPFVYSLRSPGYAI 180
DB 121 D1LRNPFCKDVIYNSKKMLKKLOESRFVDVADPIRPFCELLAEELFNPLVYSLRFTPGYIF 180
OY 181 EKISGGLLPSPSYPVVMSLSLSDMTFIERVKMIIVLYEYEFMWQIFDMKKWDQFYSEVL 240
DB 181 EKHGCGGLFPSPSYPVVMSLSLSDMTFMERVKMIIVMSLFDYQWYDMKKWDQFYSEVL 240
OY 241 GRPFTLSSETMAKADIWILIRYVWDQFPHPLIPNVEFVGLHCKRAKPLPKMESEFVSSG 300
DB 241 GRPFTLSSETMGKADIWILIRNSWNPFPHPPLIPNVEFVGLHCKRAKPLPKMESEFVSSG 300
OY 301 ENGVVFSLSGMSVNTSEERANVYASALAKI POKVLRFDGNKEDTGLNTRYLKYI PGN 360
DB 301 ENGVVFSLSGMSVNTMEERANVYASALAKI POKVLRFPQKRPDTGLNTRYLKYI PGN 360
OY 361 DLIGHPTKAFITHGNGVYEAIVYHGVNPGVPIFGDQDLDNIAMKAKGAAVEINFKTM 420
DB 361 DLIGHPTKAFITHGNGVYEAIVYHGVNPGVPIFGDQDLDNIAMKTKGAAVLDLPTM 420
OY 421 TSEDLALALRVITDSSYKKNAMLSLTIHNDQPKPLDRAVFWIEFWMRHHGAHLFSAA 480
DB 421 SSTDLNRLKTVINDPLYKEVMKLSIQHDPKPLDRAVFWIEFWMRHHGAHLRPA 480
OY 481 HDLTFPQVYSIDVIGFLTCAVATIFPTKCFELSPCKFNKTRKIERE 529
DB 481 HDLTFPQVHSLDVIGFLLCATVITFIMKCCFLCPMKFPARKKGGKSD 529

RESULT 5
UDBB_HUMAN STANDARD; PRT; 529 AA.
ID UDBB_HUMAN STANDARD; PRT; 529 AA.
AC 075310.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UDP-glucuronosyltransferase 2B11 precursor, microsomal (EC 2.4.1.17)
DE (UDPglc*).
DE UGT2B11.
GN GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98340847; Pubmed=9675083;
RA Beaulieu M., Levesque B., Hum D.W., Belanger A.;
RT "Isolation and characterization of a human orphan UDP-
RT glucuronosyltransferase, UGT2B11.";
RL Biochem. Biophys. Res. Commun. 248:44-50 (1998).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: Widely expressed.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AF016492; AAC27891.1; --
CC Genew; HGNC:12545; UGT2B11.
DR

```

DR MIN, 603064; -.
 DR GO; GO:0008210; P:estrogen metabolism; TAS.
 DR GO; GO:000805; P:xenobiotic metabolism; TAS.
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam: PF00201; UDPGT, 1.
 DR PROSITE; PS00375; UDPGT, 1.
 DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KM Multigene family; Microsome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 2 529 UDP-GLUCURONOSYLTRANSFERASE 2B11.
 FT TRANSMEM 493 513 POTENTIAL.
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 529 AA; 61038 MW; CE4AC3C71CFC2AB4 CRC64;
 Query Match 81.9%; Score 2292; DB 1; Length 529;
 Best Local Similarity 80.2%; Pred. No. 5.6e-165;
 Matches 424; Conservative 43; Mismatches 62; Indels 0; Gaps 0;
 QY 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFSHMNIKITIIDELOYRGHEVTVLSSA 60
 DB 1 MTKMTSVLLIHLSCYSSGCGKVLVWMAEYSHMMNMKITILKELOYRGHEVTVLSSA 60
 QY 61 SISDPNSPTLKEVYVSVLTKTEFEDIKQLYKMAELPKDTFWSYFSQVOSIMWTFN 120
 DB 61 SILFDPNDASTLKEVYPTSLKTEFENIMQYKRWSDIRKDSFWLTFSSOEILWELY 120
 QY 121 DILKPKCDIVSNKKMLKQESRFDVLAADVPPGELLAEILKIPVYSLRSPGYAI 180
 DB 121 DIFNPKCDIVSNKKMKQESRFDVLAADVPPGELLAEILKIPVYSLRSPGYAI 180
 QY 181 EKHSGLLPSPSYVYVVMSELSQMTFIERVKNNIYVLYFEFVFOIPDMKKDQFYSEVL 240
 DB 181 ERHSGGLFPSPSYPIVYVMSKLSQMTFIERVKNNIYVLYFEFVFOIMSKMDQFYSEVL 240
 QY 241 GRPPTLSEMAKADIMLRNWDQFPHPLPNVEFVGLHCKPAKPLPKMEEFVSSG 300
 DB 241 GRPPTLSEMAKADIMLRNWDQFPHPLPNVEFVGLHCKPAKPLPKMEEFVSSG 300
 QY 301 ENGVAVSLGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWI 360
 DB 301 ENGVAVSLGSMVNTSEERANVIATATAKIPQVLMRFPDGNKPTDGLNTRLYKWI 360
 QY 361 DLLGHPKTAFTTGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFTM 420
 DB 361 DLLGHPKTAFTTGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFTM 420
 QY 421 TSDDLRLALRTVITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFVNRHGAHLSSAA 480
 DB 421 SSTDLRLALRTVINDPSYKENIMKLSRIHQDPVKPLDRAVFWIEFVNRHGAHLSSAA 480
 QY 481 HDLTFWFGHYSTIDVIGLLTCAVTAIFLTGKFLFSCQFNKTRIKERE 529
 DB 481 HDLTFWFGHYSTIDVIGLLTCAVTAIFLTGKFLFSCQFNKTRIKERE 529
 RESULT 6
 ID UDBA_HUMAN STANDARD; PRT; 528 AA.
 AC P36537;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glucuronosyltransferase 2B10 precursor, microsomal (EC 2.4.1.17)
 DE (UDPGT).
 GN UGT2B10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=93326164; PubMed=8333863;

RA Jin C.-J., Miners J.O., Lillywhite K.J., McKenzie P.I.;
 RT "cDNA cloning and expression of two new members of the human liver
 FT UDP-glucuronosyltransferase 2B subfamily".
 RL Biochem. Biophys. Res. Commun. 194:496-503(1993).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63359; CAA44961.1; -.
 DR PIR; JN0620; JN0620.
 DR Genew; HGNC:12544; UGT2B10.
 DR MIN; 600070; -.
 DR GO; GO:0006629; P:liver metabolism; TAS.
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT, 1.
 DR PROSITE; PS00375; UDPGT, 1.
 DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KM Multigene family; Microsome.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT CHAIN 24 528 UDP-GLUCURONOSYLTRANSFERASE 2B10.
 FT TRANSMEM 492 512 POTENTIAL.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 528 AA; 60774 MW; 5B6EA3DBC032C2B0 CRC64;
 Query Match 81.5%; Score 2281.5; DB 1; Length 528;
 Best Local Similarity 80.5%; Pred. No. 3.5e-165;
 Matches 426; Conservative 44; Mismatches 58; Indels 1; Gaps 1;
 QY 1 MSMKWTSAALLIOLSCYFSSGCGKVLVWPTFSHMNIKITIIDELOYRGHEVTVLSSA 60
 DB 1 MALKMTT-VLLIOLSFSSGCGKVLVMAEYSLMMNMKITILKELOYRGHEVTVLSSA 59
 QY 61 SISDPNSPTLKEVYVSVLTKTEFEDIKQLYKMAELPKDTFWSYFSQVOSIMWTFN 120
 DB 61 SILFDPNDASTLKEVYPTSLKTEFENIMQYKRWSDIRKDSFWLTFSSOEILWAIN 119
 QY 121 DILKPKCDIVSNKKMLKQESRFDVLAADVPPGELLAEILKIPVYSLRSPGYAI 180
 DB 121 DILKPKCDIVSNKKMLKQESRFDVLAADVPPGELLAEILKIPVYSLRSPGYAI 180
 QY 181 EKHSGLLPSPSYVYVVMSELSQMTFIERVKNNIYVLYFEFVFOIPDMKKDQFYSEVL 240
 DB 181 ERHSGGLFPSPSYPIVYVMSKLSQMTFIERVKNNIYVLYFEFVFOIPDMKKDQFYSEVL 239
 QY 241 GRPPTLSEMAKADIMLRNWDQFPHPLPNVEFVGLHCKPAKPLPKMEEFVSSG 300
 DB 240 GRPPTLSEMAKADIMLRNWDQFPHPLPNVEFVGLHCKPAKPLPKMEEFVSSG 299
 QY 301 ENGVAVSLGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWI 360
 DB 300 ENGVAVSLGSMVNTSEERANVIATATAKIPQVLMRFPDGNKPTDGLNTRLYKWI 359
 QY 361 DLLGHPKTAFTTGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFTM 420
 DB 360 DLLGHPKTAFTTGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFTM 419
 QY 421 TSDDLRLALRTVITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFVNRHGAHLSSAA 480
 DB 420 SSTDLRLALRTVINDPSYKENIMKLSRIHQDPVKPLDRAVFWIEFVNRHGAHLSSAA 479

OY 481 HDLTFPHOHSIDVIGFLTYCATAIFLFTKCFLSCCKFNTRKIERE 529
 DB 480 HNLTFPHOHSIDVIGFLTYCATAIFLFTKCFLSCCKFNTRKIERE 528

RESULT 7

UDBS_HUMAN STANDARD; PRT; 529 AA.
 AC Q9BY64; Q9BY62; Q9BY63;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE UDP-glucuronosyltransferase 2B28 precursor, microsomal (EC 2.4.1.17).
 GN UGT2B28.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
 RX MEDLINE=21197745; PubMed=11300766;
 RA Levesque E., Turgeon D., Carrier J.-S., Montminy V., Beaulieu M.,
 RA Belanger A.;
 RT "Isolation and characterization of the UGT2B28 cDNA encoding a novel
 RT human steroid conjugating UDP-glucuronosyltransferase.";
 RT Biochemistry 40:3869-3881(2001).
 CC -1- FUNCTION: UDPGTs are of major importance in the conjugation and
 CC subsequent elimination of potentially toxic xenobiotics and
 CC endogenous compounds. This isozyme has glucuronidating capacity
 CC with steroid substrates such as 5-beta-androstane 3-alpha,17-beta-
 CC diol, estradiol, ADR, eugenol and bile acids. Only isoform 1 seems
 CC to be active.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1; Synonyms=1;
 CC IsoId=Q9BY64-1; Sequence=Displayed;
 CC Name=2; Synonyms=1;
 CC IsoId=Q9BY64-2; Sequence=VSP_006710, VSP_006711;
 CC Name=3; Synonyms=1;
 CC IsoId=Q9BY64-3; Sequence=VSP_006712;
 CC -1- TISSUE SPECIFICITY: Expressed in the liver, breast and kidney.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF177272; AAK31807.1; -;
 CC EMBL; AF177272; AAK31808.1; -;
 CC EMBL; AF177274; AAK31809.1; -;
 CC GeneW; HGNC:13479; UGT2B28.
 CC MIM; 606497; -;
 CC GO; GO:0005783; C:cytoplasmic reticulum; IDA.
 CC GO; GO:0006805; P:xenobiotic metabolism; IDA.
 CC InterPro; IPR002213; UDP_gluco_trans.
 CC Pfam; PF00201; UDPGT; 1.
 CC PROSITE; PS00375; UDPGT; 1.
 CC Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 CC Multigene family; Microsome; Alternative splicing.
 CC SIGNAL
 CC CHAIN 1 24
 CC TRANSMEM 25 529
 CC CARBOHYD 315 517
 CC VARSPLIC 335 335
 CC FTID=VSP_006710.
 CC /FTID=VSP_006710.

FT VARSPLIC 336 529 Missing (in isoform 2).
 FT FTID=VSP_006711.
 FT VARSPLIC 105 221 Missing (in isoform 3).
 FT FTID=VSP_006712.
 SO SEQUENCE 529 AA; 60906 MW; 8C75277E564690C1 CRC64;

Query Match 80.5%; Score 2253; DB 1; Length 529;
 Best Local Similarity 79.0%; Pred. No. 5e-163;
 Matches 416; Conservative 45; Mismatches 66; Indels 0; Gaps 0;

OY 1 MSMTGTAALLIQLSCYFSSGCKKLVLPFEFHHNNITLDELVQHGHEVLAASA 60
 DB 1 MALMTGTVLLIHIGCYFSSGCKKLVWGEVHHNNITLDELVQHGHEVLAASA 60
 OY SIFSDPNSPTLKEPVVPSLTKTEFEDIIKOLVKRAEALPKOTFWGVSQVQGEIMWTFN 120
 DB 61 SLDPDPAFLTKLEVPVPSLTKTEFENIIMQVKRSDIOKDSFWLYFSQGEIIMWTFN 120
 OY 121 DILRPCKDIYSNKKMKKQESRFDVLDVAVPFGELLAELLKIPVYSLRSPGYAI 180
 DB 121 DIFNPKCDVSNKKMKKQESRFDIIPDAFPCEGLLAALLNIPVYSLSCTPGYTI 180
 OY 181 EKHSGGLFPSPSYVPMVMSLSDMTFLERKMMIYVLFYEFWFQIDMKKKDQFYSVL 240
 DB 181 ERHSGGLIFPSPSYVPMVMSKLSDMTEMRKMMIYVLFYEFWFQIDMKKKDQFYSVL 240
 OY 241 GRPTLSEPMKADIMLIRNWDQFPPLPENVFVGLGCKCPAKPLPKMEEFVQSSG 300
 DB 241 GRPTLSEPMKADIMLIRNWSQFPPLPENVFVGLGCKCPAKPLPKMEEFVQSSG 300
 OY 301 ENGVSFSLGSMVNTSEERANVIAALAKIPQVTLRFPDGNKPDGTLNTRYKWIPON 360
 DB 301 ENGVSFSLGSMVNTSEERANVIAALAKIPQVTLRFPDGNKPDGTLNTRYKWIPON 360
 OY 361 DLHPKTKAFTTGGMNGIYEAALYGVPMGVIPFGDQDNLNHRKAGAAVEINFTM 420
 DB 361 DLHPKTKAFTTGGMNGIYEAALYGVPMGVIPFGDQDNLNHRKAGAAVEINFTM 420
 OY 421 TSEDLLRLRTVIVDSYKEMAKLSRIHQDPKPLDRAVFIETVNRHKGAKLSAA 480
 DB 421 TSEDLLRLRTVIVDSYKEMAKLSRIHQDPKPLDRAVFIETVNRHKGAKLSAA 480
 OY 481 HDLTFPHOHSIDVIGFLTYCATAIFLFTKCFLSCCKFNTRKIERE 529
 DB 481 HDLTFPHOHSIDVIGFLTYCATAIFLFTKCFLSCCKFNTRKIERE 529

RESULT 8

UDBS_HUMAN STANDARD; PRT; 530 AA.
 AC 075735;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE UDP-glucuronosyltransferase 2B17 precursor, microsomal (EC 2.4.1.17)
 GN (UDPGR) (C19-steroid specific UDP-glucuronosyltransferase).
 GN UGT2B17.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC TISSUE: Prostate;
 RX MEDLINE=96394358; PubMed=8798464;
 RX Beaulieu M., Levesque E., Hum D.W., Belanger A.;
 RT "Isolation and characterization of a novel cDNA encoding a human
 RT UDP-glucuronosyltransferase active on C19 steroids.";
 RT J. Biol. Chem. 271:22855-22862(1996).
 RL (2)
 RP SEQUENCE FROM N. A.
 RP MEDLINE=98030207; PubMed=9364925;
 RX Beaulieu M., Tchernof A., Beatty B.G., Belanger A.,

RA Hum D.W.;
 RT "Chromosomal localization, structure, and regulation of the UGT2B17
 RT gene, encoding a C19 steroid metabolizing enzyme.";
 RL DNA Cell Biol. 16:1143-1154(1997).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS. THE MAJOR SUBSTRATES OF THIS ISOZYME ARE
 CC EUGENOL > 4-METHYLBELLEREPONE > DIHYDROTESTOSTERONE (DHT) >
 CC ANDROSTANE-3ALPHA,17BETA-DIOL (3ALPHA-DIOL) > TESTOSTERONE >
 CC ANDROSTERONE (ADT).
 CC CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC TISSUE SPECIFICITY: EXPRESSED IN VARIOUS TISSUES INCLUDING THE
 CC LIVER, KIDNEY, TESTIS, UTERUS, PLACENTA, MAMMARY GLAND, ADRENAL
 CC GLAND, SKIN, AND PROSTATE.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: U59209; AAC25491.1; -
 CC Genew; HGNC:12547; UGT2B17.
 CC DR MIM; 601903; -
 CC DR GO; GO:0005624; C:membrane fraction; TAS.
 CC DR GO; GO:0008202; P:steroid metabolism; TAS.
 CC DR InterPro: IPR002213; UDP_gluco_trans.
 CC DR Pfam: PF00201; UDPGT; 1.
 CC DR PROSITE; PS00375; UDPGT; 1.
 CC KM Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 CC MultiGene family; Microsome.
 CC FT SIGNAL 1 23 POTENTIAL.
 CC FT CHAIN 1 24 530 UDP-GLUCURONOSYLTRANSFERASE 2B17.
 CC FT TRANSMEM 495 515 POTENTIAL.
 CC FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 530 AA; 61095 MW; 8E59EBCA3CFA3760 CRC64;
 CC
 CC Query Match 77.5%; Score 2169.5; DB 1; Length 530;
 CC Best Local Similarity 74.7%; Pred. No. 11e-156;
 CC Matches 396; Conservative 60; Mismatches 73; Indels 1; Gaps 1;
 CC
 CC QY 1 MSMKWTSLALLIOLSCYPSGSCGKVLVWPFESHMNIKITILDELVORGHVTVLASSA 60
 CC Db 1 MSLKMSVFLMLQSLSCFSSGCGKVLVWPFESHMNIKITILDELVORGHVTVLASSA 60
 CC QY 61 SISFDPSPTLKEFVYVPSLTKEFEDIIKOLYKMA-ELPKDTFMSYFSQVOEIMWTF 119
 CC Db 61 SILVNASKSAIKLEVPYTSLTKNLDEDFPMKMGDMRTYTSISKTFMSYFQQLDELCEY 120
 CC QY 120 NDILRKCKDLYSNKKMLKQESRPDYVLADANVPPELLAELIKLPFVSLRSPRYA 179
 CC Db 121 SDVYIKCEDAVLNKKMRKIQESKFDVLADANVPCCELLAEILNLPFLYSLSRFSYGT 180
 CC QY 180 IEKSGGLPPSPVVPVMSLSDOMTIERVKNNIYLYPEPFQJFDMKKMQOFSYEV 239
 CC Db 181 VERKGGGLPPSPVVPVMSLSDOMTIERVKNNIYLYPEPFQJFDMKKMQOFSYEV 240
 CC QY 240 LGRPTTSETMAKADIMLIRNYWDFQPHPLLPVNEFVGLAHCPAKPLKEMEEFYQSS 299
 CC Db 241 LGRPTTLETWGAEMWLIRTYWDEBPFRPLPVNDVFGALHCPAKPLKEMEEFYQSS 300
 CC QY 300 GENGVVFSLGSMSNTSEEPANVYASALAKIPQKVLMPFGNKPRTGLNTRLYKMPQ 359
 CC Db 301 GENGIIVFSLGSMISNSESANNIASALAOIPQVLMPFGNKPRTGLNTRLYKMPQ 360
 CC QY 360 NDILGHPRTKAFITHGNGIYEALYHGVPVWGVIPFGDQDNIAHMKAKGAIVEINPKT 419

Db 361 NDILGHPRTKAFITHGNGIYEALYHGVPVWGVIPFGDQDNIAHMKAKGAALSDVIRT 420
 QY 420 MTSDDLRLARTYITDSYKENMRSLRHDDQPVKRLDAVPIEVMHKKAKHLRSA 479
 Db 421 MSSDDLNLAKSYINDPIYENKMLSRHHDDQPVKRLDAVPIEVMHKKAKHLRVA 480
 QY 480 AHDLPWFQHSIDVIGFLVCVATAPLPTKCFAPSCQKFNKRIKRE 529
 Db 481 AHDLPWFQHSIDVIGFLVCVATAPLPTKCFAPSCQKFNKRIKRE 530
 CC
 CC RESULT 9
 CC UDBF HUMAN STANDARD; PRT; 530 AA.
 CC ID UDBF HUMAN
 CC AC P54855; P23765; Q9URK63;
 CC DT 01-NOV-1991 (Rel. 20, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE UDP-glucuronosyltransferase 2B15 precursor, microsomal (EC 2.4.1.17)
 CC (UDPGT) (UDPGTH-3) (HLUG4).
 CC GN UGT2B15 OR UGT2B8.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Liver;
 CC RX MEDLINE=95136867; PubMed=7835232;
 CC RA Green M.D., Oltu E.M., Tepfly T.R.;
 CC RT "Stable expression of a human liver UDP-glucuronosyltransferase
 CC (UGT2B15) with activity toward steroid and xenobiotic substrates.";
 CC RL Drug Metab. Dispos. 22:799-805(1994).
 CC RN [2]
 CC RP SEQUENCE OF 8-530 FROM N.A., AND PARTIAL SEQUENCE.
 CC RC TISSUE=Liver;
 CC RX MEDLINE=90343358; PubMed=2116769;
 CC RA Coffman B.L., Tepfly T.R., Irshaid Y.M., Green M.D., Smith C.,
 CC Jackson M.R., Wooster R., Burchell B.;
 CC RT "Characterization and primary sequence of a human hepatic microsomal
 CC estriol UDPglucuronosyltransferase.";
 CC RL Arch. Biochem. Biophys. 281:170-175(1990).
 CC RN [3]
 CC RP SEQUENCE FROM N.A. AND VARIANT TYR-85.
 CC RX MEDLINE=97439504; PubMed=9295060;
 CC RA Levesque E., Beaulieu M., Green M.D., Tepfly T.R., Belanger A.,
 CC Hum D.W.;
 CC RT "Isolation and characterization of UGT2B15(Y85): a
 CC UDP-glucuronosyltransferase encoded by a polymorphic gene.";
 CC RL Pharmacogenetics 7:317-325(1997).
 CC RN [4]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Liver;
 CC RX Owens I.S.;
 CC RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC CC -1- FUNCTION: UDPGTs are of major importance in the conjugation and
 CC subsequent elimination of potentially toxic xenobiotics and
 CC endogenous compounds. This isozyme displays activity toward
 CC several classes of xenobiotic substrates, including simple
 CC phenolic compounds, 7-hydroxylated coumarins, flavonoids,
 CC anthraquinones, and certain drugs and their hydroxylated
 CC metabolites. It also catalyzes the glucuronidation of endogenous
 CC estrogens and androgens.
 CC CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC CC -1- TISSUE SPECIFICITY: Expressed in many tissues.
 CC CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC CC -1- CAUTION: REF. 2 THOUGHT THAT THIS WAS A SEPARATE FORM (UGT2B8). THE
 CC NAME UGT2B8 IS NOW BEING REUSED FOR A RAT ENZYME.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U08854; AAC50077.1; -
 DR EMBL: U06641; AAA83406.1; -
 DR EMBL: AF180322; AAD55093.1; -
 DR EMBL: AF548389; AAN40695.1; -
 DR PIR: A48633; A48633.
 DR Genew: HGNC:12546; UGT2B15.
 DR MIM: 600069; -
 DR GO: GO:0008202; P:steroid metabolism; TAS.
 DR GO: GO:0006805; P:xenobiotic metabolism; TAS.
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam: PF00201; UDPGT; 1.
 DR PROSITE: PS00375; UDPGT; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KW Multigene family; Microsome; Polymorphism.
 FT SIGNAL 1 23
 FT CHAIN 1 23
 FT TRANSMEM 24 530
 FT CARBOHYD 495 515
 FT CARBOHYD 65 65
 FT CARBOHYD 316 316
 FT CARBOHYD 483 483
 FT VARIANT 85 85
 FT FT
 FT CONFLICT 119 119
 FT CONFLICT 145 145
 FT CONFLICT 150 155
 FT CONFLICT 162 162
 FT CONFLICT 165 165
 FT CONFLICT 170 177
 FT CONFLICT 181 203
 FT CONFLICT 203 203
 FT CONFLICT 293 293
 FT CONFLICT 401 401
 FT CONFLICT 443 443
 FT CONFLICT 501 501
 FT CONFLICT 523 523
 FT FT
 SQ SEQUENCE 530 AA; 60961 MW; 23DDA4B9687CB4 CRC64;
 Query Match 77.1%; Score 2158.5; DB 1; Length 530;
 Best Local Similarity 74.5%; Pred. No. 7,2e-156;
 Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;
 Oy 1 MSMKWTALLILQLSCYFSSGCGKVLWPTFESHMMNIKTILDELVORGEVTVLAASA 60
 Db 1 MSIKWTYPLLIQLSCYFSSGCGKVLWPTFESHMMNIKTILDELVORGEVTVLAASA 60
 Oy 61 SISDPNPSPTLKFEVYVSLTKTEFEDI IKQLVKRA-ELPKDTFMSYFSGVOEIMWTF 119
 Db 61 STLVASXSAAIKLEVYPTSLTKNLEDSLKILRLRWYVSGKTFWFSFQSLQELCWEX 120
 Oy 120 NDILAKFKCDVYNNKKMKKIOESFEDVVLDAVPPPELLAELIKIPVYSIRSPGYA 179
 Db 121 YDYSNKLCKDAVLNKKMKKIOESFEDVVLDAVPPPELLAELIKIPVYSIRSPGYA 180
 Oy 180 IEKHSGLLFPSPYVVMSELDQMTFIERVKNKIYVLFEPFQIFDMKMDQFYSEV 239
 Db 181 FEKNGGGLFPSPYVVMSELDQMTFIERVKNKIYVLFEPFQIFDMKMDQFYSEV 240
 Oy 240 LGRPTTLEETMAKADIMLRINWDFQPHPLLPNVEFVGLHCKCAKPLPKEMEEFYOSS 299
 Db 241 LGRPTTLEETMAKADIMLRINWDFQPHPLLPNVEFVGLHCKCAKPLPKEMEEFYOSS 300
 Oy 300 GENGIVVSLGSMVNTSEERANVTASALAKIPQVLRPFQDNKDDTGLNTRLYKWTIPQ 359
 Db 301 GENGIVVSLGSMVNTSEERANVTASALAKIPQVLRPFQDNKDDTGLNTRLYKWTIPQ 360
 Oy 360 NDLLGHPTKAPITHTGNMGIYEAIVHGVPMVGPVIFGDQDNLINAMKAKGAIVEINFKT 419

Db 361 NDLLGHPTKAPITHTGNMGIYEAIVHGVPMVGPVIFGDQDNLINAMKAKGAALSVIRT 420
 Oy 420 MTSDDLRLARTVTTDSYSENNMRLSRHHDDQVRKLDRAVFIETVMRHKGAKHLRSA 479
 Db 421 MSSDDLNLAKSVINDVYENVKLSRIHDDQMKPLDRAVFIETVMRHKGAKHLRVA 480
 Oy 480 AHDLTWFQHSIDVIGFLNCVATAIFLFTKCFPLSCQKFNKTKIEKRE 529
 Db 481 AHNLTWIOHSLDVIAFLACVATVIFITKFCIFCRKLAKTKOKKKKD 530

RESULT 10
 UDBK MACFA
 ID UDBK MACFA STANDARD; PRT; 530 AA.
 AC 077649;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glycuronosyltransferase 2B20 precursor, microsomal (EC 2.4.1.17)
 DE (UNPRT).
 GN UGT2B20.
 OS Macaca faecicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC TISSUE=Liver, and Prostate;
 RX MEDLINE=99112924; PubMed=9895303;
 RA Barbier O., Belanger A., Hum D.W.;
 RT "Cloning and characterization of a simian UDP-glycuronosyltransferase
 enzyme UGT2B20, a novel C19 steroid-conjugating protein.";
 RL Biochem. J. 337:567-574(1999).
 CC -1- FUNCTION: UDPGs are of major importance in the conjugation and
 CC subsequent elimination of potentially toxic xenobiotics and
 CC endogenous compounds. This isozyme has glucuronidating capacity
 CC with androgens, such as testosterone, dihydrotestosterone (DHT)
 CC and 3alpha-diol. It is also active on catecholoeestrogens including
 CC 1,3,5,10-oestradiene-3,4-diol-17-one.
 CC -1- CATALYTIC ACTIVITY: UDP-glycuronosyl + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Query Match 75.1%; Score 2100.5; DB 1; Length 530;
 Best Local Similarity 73.1%; Pred. No. 1.8e-151;
 Matches 386; Conservative 60; Mismatches 81; Indels 1; Gaps 1;

```

QY 1 MSKMTSALLIOLSCYFSSGCGKVLVWPTFSHMNIKTILDELVQGHVTLASSA 60
DB 1 MSKMTSVFLLIOLSCYFSSGCGKVLVWPTFSHMNIKTILDELVQGHVTLASSA 60
QY 61 SISDPNSPTLKEVPVSLTKTEFEDIIKOLYKMA-ELPKDTFMSYSQVOEIMWF 119
DB 61 STFNVDKSSSAIKFEVPTSLTKMDMEDSLMKLIDITWYSISNFTPLSYESKLOELCMEY 120
QY 120 NDILRKFCXKIVSKMKLQESRFPVVLADAVFPFGEILLAEELKIPFYSLRFSQGYA 179
DB 121 YYISEKCKOAVLKKMTLTKETKFTVILLADALNPGELAELEFNIIPYISLFTVGYT 180
QY 180 IEKSGGLPEPSPVYVPMSELSDQMTFIERVKMIVLYFEFQOIFDMKMDQFYSEV 239
DB 181 FEKNGGFLPPSPVPMSELSDQMTFIERVKMIVLYFEFQOIFDMKMDQFYSEV 240
QY 240 LGRPTTSETMAKADIVLIRNYWDFOPPHLLPNVEVYGLHCKPAVLPKMEEFVQSS 299
DB 241 LGRPTTSETMAKADIVLIRNYWDFOPPHLLPNVEVYGLHCKPAVLPKMEEFVQSS 300
QY 300 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRLYKMIPO 359
DB 301 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRLYKMIPO 360
QY 360 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPFGDQLDNIJAHKAKAAVEINFKT 419
DB 361 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPFGDQLDNIJAHKAKAAVEINFKT 420
QY 420 MTSDDLRLALRTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMMHKGAKHLRSA 479
DB 421 MSSDDLRLALRTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMMHKGAKHLRSA 480
QY 480 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFPSQCKNKRKTRE 527
DB 481 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFPSQCKNKRKTRE 528

RESULT 11
-----
ID _UNDE_RABIT STANDARD; PRT; 530 AA.
AC P36513;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B14 precursor, microsomal (EC 2.4.1.17)
DE (UDPgt) (EGT12).
GN UGT2B14.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=93315511; Pubmed=8325897;
RA Turkey R.H., Pendurthi U.R., Nguyen N.T., Green M.D., Teply T.R.;
RT "Cloning and characterization of rabbit liver UDP-
RT glucuronosyltransferase cDNAs. Developmental and inducible expression
RT of 4-hydroxydiphenyl UGT2B13."
RL J. Biol. Chem. 268:15260-15266(1993).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDOGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN ADULT RABBITS.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: L01082; AAA18021.1; -.
CC PIR: C47113; C47113.
CC InterPro: IPR002213; UDP_gluco_trans.
CC Pfam: PF00201; UDPGT; 1.
CC PROSITE: PS00375; UDPGT; 1.
CC TRANSFAS: Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
CC Multigene family; Microsome.
CC SIGNAL 24
CC CHAIN 25 530 UDP-GLUCURONOSYLTRANSFERASE 2B14.
CC TRANSMEM 494 510 POTENTIAL.
CC CARBOHYD 134 134 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 316 316 N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 530 AA; 60664 MW; CEAA4BF25B53CB5 CRC64;

Query Match 71.2%; Score 1991.5; DB 1; Length 530;
Best Local Similarity 67.9%; Pred. No. 3,2e-143;
Matches 360; Conservative 79; Mismatches 90; Indels 1; Gaps 1;

QY 1 MSKMTSALLIOLSCYFSSGCGKVLVWPTFSHMNIKTILDELVQGHVTLASSA 59
DB 1 MSKMTSVFLLIOLSCYFSSGCGKVLVWPTFSHMNIKTILDELVQGHVTLASSA 60
QY 60 ASISFPNSPTLKEVPVSLTKTEFEDIIKOLYKMAELPKDTPMSYSQVOEIMWF 119
DB 61 ASISFPNSPTLKEVPVSLTKTEFEDIIKOLYKMAELPKDTPMSYSQVOEIMWF 120
QY 120 NDILRKFCXKIVSKMKLQESRFPVVLADAVFPFGEILLAEELKIPFYSLRFSQGYA 179
DB 121 SDCENAKCEVAVNKLMTLQESRFDILSDAIGCGELLAELKIPFYSLRFTTGYT 180
QY 180 IEKSGGLPEPSPVYVPMSELSDQMTFIERVKMIVLYFEFQOIFDMKMDQFYSEV 239
DB 181 MEKSGGLPEPSPVYVPMSELSDQMTFIERVKMIVLYFEFQOIFDMKMDQFYSEV 240
QY 240 LGRPTTSETMAKADIVLIRNYWDFOPPHLLPNVEVYGLHCKPAVLPKMEEFVQSS 299
DB 241 LGRPTTSETMAKADIVLIRNYWDFOPPHLLPNVEVYGLHCKPAVLPKMEEFVQSS 300
QY 300 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRLYKMIPO 359
DB 301 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRLYKMIPO 360
QY 360 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPFGDQLDNIJAHKAKAAVEINFKT 419
DB 361 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPFGDQLDNIJAHKAKAAVEINFKT 420
QY 420 MTSDDLRLALRTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMMHKGAKHLRSA 479
DB 421 MSSDDLRLALRTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMMHKGAKHLRSA 480
QY 480 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFPSQCKNKRKTRE 529
DB 481 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFPSQCKNKRKTRE 530

RESULT 12
-----
ID _UNDE_RAT STANDARD; PRT; 529 AA.
AC P09875;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B1 precursor, microsomal (EC 2.4.1.17)
DE (UDPgt) (UDPGTR-2).
GN UGT2B1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=66196018; PubMed=3084479;
RA McKenzie P.I.;
RT "Rat liver UDP-glucuronosyltransferase. Sequence and expression of a
  RT CDNA encoding a phenobarbital-inducible form.";
RL J. Biol. Chem. 261:6119-6125(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90293083; PubMed=2113533;
RA McKenzie P.I.; Roddoun L.;
RT "Organization of the rat UDP-glucuronosyltransferase, UDPGT-2, gene.
  RT and characterization of its promoter.";
RL J. Biol. Chem. 265:11328-11333(1990).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
  CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
  CC ENDOGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
  CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- INDUCTION: By phenobarbital.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
DR EMBL; M13506; AAA42313.1; -
DR EMBL; M35086; AAA42310.1; JOINED.
DR EMBL; M35202; AAA42310.1; JOINED.
DR EMBL; M35080; AAA42310.1; JOINED.
DR EMBL; M35082; AAA42310.1; JOINED.
DR EMBL; M35083; AAA42310.1; JOINED.
DR PIR; A42233; A42233.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 529 UDP-GLUCURONOSYLTRANSFERASE 2B1.
FT TRANSMEM 494 510 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 529 AA; 60484 MW; 14DF0224B1C3044 CRC64;

Query Match 70.5%; Score 1973.5; DB 1; Length 529;
Best Local Similarity 68.4%; Pred. No. 7.3e-142;
Matches 361; Conservative 68; Mismatches 98; Indels 1; Gaps 1;

```

```

Db 241 LGRPTLTTEMKADILWIRTFWDLFFPHPLPNFDFVGLHCKPAKPLFREHEEYQSS 300
Oy 300 GENGWVVFSLGSMVSNTESEERANVIASALAKIPQKVLMPEDGNKPDITGLNTRLYKMIPO 359
Db 301 GENGWVVFSLGSMVSNTESEERANVIASALAKIPQKVLMPEDGNKPDITGLNTRLYKMIPO 360
Oy 360 NDILGHPKTAFTTHGGMNIGYEAIIYGVPMGVPIFGDLDNIAMKAGAVEINFKT 419
Db 361 NDILGHPKTAFTTHGGMNIGYEAIIYGVPMGVPIFGDLDNIAMKAGAVEINFKT 420
Oy 420 MTSDDLRAIRTYITDSSYKENMRSLRIHDDQVRELDRAVFEFVMEHKGAKHLRSA 479
Db 421 LSTGLLTAKIYWNDSYENMRSLRIHDDQVRELDRAVFEFVMEHKGAKHLRST 480
Oy 480 AHDLTWFQHSIDYIGFLTCVATAPLFTKCFPLSCOKENKTRKIEK 527
Db 481 LHDLSWFQHSIDYIGFLTCVATAPLFTKCFPLSCOKENKTRKIEK 528

```

```

RESULT 13
ID UDPG_RABBIT STANDARD; PRT; 523 AA.
AC 019103;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B16 precursor, microsomal (EC 2.4.1.17)
DE (UNPRT) (Fragment).
GN UGT2B16.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97166166; PubMed=9013565;
RA Li Q., Lou X., Peyronneau M.-A., Straub P.O., Tukey R.H.;
RT "Expression and functional domains of rabbit liver UDP-
  RT glucuronosyltransferase 2B16 and 2B13.";
RL J. Biol. Chem. 272:3272-3279(1997).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
  CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
  CC ENDOGENOUS COMPOUNDS. ACTS ON SMALL PHENOLIC AGENTS SUCH AS 2-
  CC NAPHTHOL AND 4-METHYLBUMBELIFERONE AS WELL AS BULKY PHENOLIC
  CC COMPOUNDS LIKE 2-HYDROXY- AND 4-HYDROXYBIPHENYL. IN CONTRAST TO
  CC 2B13 IT IS ACTIVE TOWARD 4-HYDROXYESTERONE.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
  CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN ADULT RABBITS.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
DR EMBL; U72742; AAB71494.1; -
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome.
FT SIGNAL 1 16
FT CHAIN 17 523 UDP-GLUCURONOSYLTRANSFERASE 2B16.
FT TRANSMEM 487 503 POTENTIAL.
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 523 AA; 60077 MW; 22408BE768ED3DE CRC64;

```

Query Match 69.3%; Score 1938.5; DB 1; Length 523;
 Best Local Similarity 68.6%; Pred. No. 3.2e-139;
 Matches 358; Conservative 64; Mismatches 99; Indels 1; Gaps 1;

QY 9 LLLIOLSCYFSSGSGKYLWPTFESHMMNITITIDELVQREHETVLASSISFPDPS 68
 |||||
 DB 2 LLLIOLSCYFSSGSGKYLWPTFESHMMNITITIDALVQREHETVLASSISFPDPS 61
 |||||

QY 69 PSTLKEEYVPVSLTTEFEEDIKQ-LVKKMAELPDTFMSYSPQVQEIIMTNDILKREK 127
 |||||
 DB 62 ESGITFEFPPTTSTDEMEAFPMYMLNKLINDVSDALMEFQTOQKPFMEISDYENIC 121
 |||||

QY 128 KOIVSNKKMLKQESRPDVLADVPFPGELAEELKIPVYSLRFSFGAIEKHSGL 187
 |||||
 DB 122 KDVLNKKIMAKLQSRFDVLADIPACGELLAEELKIPVYSLRFSFGAIEKHSGL 181
 |||||

QY 188 LFPPEYVYVMSSELSQMTFIERVKMIVLYLFEFVFOIFDMKKDQYSEVLAGPTTLLS 247
 |||||
 DB 182 LFPPEYVYVMSSELSQMTFIERVKMIVLYLFEFVFOIFDMKKDQYSEVLAGPTTLLS 241
 |||||

QY 248 ETMAKADIMLIRNMYDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSGEYVVF 307
 |||||
 DB 242 ELVGAELIMLIRSYDLEFPRPLPNVSYFGGLHCKPAKPLPKEMEFPVQSGEYVVF 301
 |||||

QY 308 SLGSMVSTSEBRANVIALAKIPQKYLMPFDGKRPDTLGIATRLYKMIPONDILGHK 367
 |||||
 DB 302 SLGSMVSTSEBRANVIALAKIPQKYLMPFDGKRPDTLGIATRLYKMIPONDILGHK 361
 |||||

QY 368 TKAFITTHGGMGIYALIHGVPMGVPIFGDOLNIAHMKAGAAVEINFTMTSEDLIR 427
 |||||
 DB 362 SKAFITTHGGMGIYALIHGVPMGVPIFGDOLNIAHMKAGAAVEINFTMTSEDLIR 421
 |||||

QY 428 ALRTYTOSYKXENMRLSRIHHDQVPLDRAVWIEFVWHRKAKHLSAHDLMTEQ 487
 |||||
 DB 422 ALRTYTOSYKXENMRLSRIHHDQVPLDRAVWIEFVWHRKAKHLSAHDLMTEQ 481
 |||||

QY 488 HYSIDVIGFLTCVATAIFLFTKCFPFSCQKFNKTRKIEKTE 529
 |||||
 DB 482 HYSIDVIGFLTCVATAIFLFTKCFPFSCQKFNKTRKIEKTE 523
 |||||

RESULT 14
 ID UDBD_RABIT STANDARD; PRT; 531 AA.
 AC P36512;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DB UDP-glucuronosyltransferase 2B13 precursor, microsomal (EC 2.4.1.17)
 DE (UDPgt) (EGT10).
 GN UGT2B13.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Liver;
 RX MEDLINE=93315511; PubMed=8325897;
 RA Tuley R.H., Penderguth U.R., Nguyen N.T., Green M.D., Tephly T.R.;
 RT "Cloning and characterization of rabbit liver UDP-
 RT glucuronosyltransferase cDNAs. Developmental and inducible expression
 RT of 4-hydroxybiphenyl UGT2B13".
 RL J. Biol. Chem. 268:15260-15266(1993).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS. ACTS ON SMALL PHENOLIC AGENTS SUCH AS 2-
 CC NAPHTOL AND 4-METHYLMELIFERONE AS WELL AS BULKY PHENOLIC
 CC COMPOUNDS LIKE 2-HYDROXY- AND 4-HYDROXYBIPHENYL. IN CONTRAST TO
 CC 2B16 IT IS ACTIVE TOWARD OCTYDILATE.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN ADULT RABBITS.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLUCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC DR EMBL; L01081; AA018020.1; -.
 CC DR PIR; B47113; B47113.
 CC DR InterPro; IPR002213; UDP_gluco_trans.
 CC Pfam; PF00201; UDPGT. 1.
 CC DR PROSITE; PS00375; UDPGT. 1.
 CC KM Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 CC Multigene family; Microsome.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 531 UDP-GLUCURONOSYLTRANSFERASE 2B13.
 CC FT TRANSMEM 495 511 POTENTIAL (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 531 AA; 60552 MW; 961DA58AC4CB9932 CRC64;

Query Match 69.3%; Score 1938; DB 1; Length 531;
 Best Local Similarity 68.7%; Pred. No. 3.6e-139;
 Matches 365; Conservative 62; Mismatches 102; Indels 2; Gaps 2;

QY 1 MSMKMTSA-LLLIOLSCYFSSGSGKYLWPTFESHMMNITITIDELVQREHETVLASS 59
 |||||
 DB 1 MPVACISVLLLIOLSCYFSSGSGKYLWPTFESHMMNITITIDALVQREHETVLASS 60
 |||||

QY 60 ASISFPDPSBSTLKEEYVPVSLTTEFEEDIKQ-LVKKMAELPDTFMSYSPQVQEIIMT 118
 |||||
 DB 61 ASISFPDPSBSTLKEEYVPVSLTTEFEEDIKQ-LVKKMAELPDTFMSYSPQVQEIIMT 120
 |||||

QY 119 FNDILKRFCDIVSNKKMLKQESRPDVLADVPFPGELAEELKIPVYSLRFSFGY 178
 |||||
 DB 121 YSDICEDICEVILNKKMLKQESRPDVLADVPFPGELAEELKIPVYSLRFSFGY 180
 |||||

QY 179 AIEKHSGLLFPPEYVYVMSSELSQMTFIERVKMIVLYLFEFVFOIFDMKKDQYSE 238
 |||||
 DB 181 MLQKHGGGLLPPEYVYVMSSELSQMTFIERVKMIVLYLFEFVFOIFDMKKDQYSE 240
 |||||

QY 239 VLGRPTTSETMAKADIMLIRNMYDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPVQ 298
 |||||
 DB 241 VLGRPTTSETMAKADIMLIRNMYDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPVQ 300
 |||||

QY 299 SGENGVYVFSIGSNVSTSEBRANVIALAKIPQKYLMPFDGKRPDTLGIATRLYKMI 358
 |||||
 DB 301 SGENGVYVFSIGSNVSTSEBRANVIALAKIPQKYLMPFDGKRPDTLGIATRLYKMI 360
 |||||

QY 359 ONDLIGHPTKAFITTHGGMGIYALIHGVPMGVPIFGDOLNIAHMKAGAAVEINFK 418
 |||||
 DB 361 ONDLIGHPTKAFITTHGGMGIYALIHGVPMGVPIFGDOLNIAHMKAGAAVEINFK 420
 |||||

QY 419 TMTSEDLIRALRTYTOSYKXENMRLSRIHHDQVPLDRAVWIEFVWHRKAKHLSA 478
 |||||
 DB 421 TMTSEDLIRALRTYTOSYKXENMRLSRIHHDQVPLDRAVWIEFVWHRKAKHLSA 480
 |||||

QY 479 AAHDLMTEQFQKSIDVIGFLTCVATAIFLFTKCFPFSCQKFNKTRKIEKTE 529
 |||||
 DB 481 AAHDLMTEQFQKSIDVIGFLTCVATAIFLFTKCFPFSCQKFNKTRKIEKTE 531
 |||||

RESULT 15
 ID UDBD_RAT STANDARD; PRT; 530 AA.
 AC P36511;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE UDP-glucuronosyltransferase 2B12 precursor, microsomal (EC 2.4.1.17)
 DE (UDPGLT).
 GN UGT2B12.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney, and Liver;
 RX MEDLINE=96032669; PubMed=7547472;
 RA Green M.D., Clarke D.J., Ocuru E.M., Styczynski P.B.,
 RA Jackson M.R., Burchell B., Teply T.R.;
 RT "Cloning and expression of a rat liver phenobarbital-inducible UDP-
 glucuronosyltransferase (2B12) with specificity for monoterpenoid
 alcohols";
 RL Arch. Biochem. Biophys. 322:460-468(1995).
 RN [2]
 RP SEQUENCE OF 24-38, AND CHARACTERIZATION.
 RC STRAIN=Mistar; TISSUE=Liver;
 RX MEDLINE=91312227; PubMed=1906977;
 RA Styczynski P.B., Green M.S., Puig J., Coffman B.L., Teply T.R.;
 RT "Purification and properties of a rat liver phenobarbital-inducible
 4-Hydroxybiphenyl UDP-glucuronosyltransferase";
 RL Mol. Pharmacol. 40:80-84(1991).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 ENOGENOUS COMPOUNDS. CATALYZES THE TRANSFER OF GLUCURONIC ACID
 FROM UDP-GLUCURONIC ACID TO VARIOUS AGLYCONE MOLECULES. CATALYZES
 THE GLUCONIDATION OF MONOTERPENOID ALCOHOLS, SUCH AS (-)-
 BONEOL, (+)-MENTHOL, AND (-)-NOPEL. IN ADDITION, A NUMBER OF
 SIMPLE PHENOLIC COMPOUNDS, SUCH AS HYDROXYBIPHENYLS, 7-
 HYDROXYLATED COUMARINS, P-NITROPHENOL, AND FOOD-DERIVED SUBSTANCES
 (E.G., NAINEBIN AND EUGENOL), AND 4-METHYLBELLIFERONE ARE ALSO
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- TISSUE SPECIFICITY: LIVER. LOWER LEVELS SEEN IN THE KIDNEY AND
 TESTIS.
 CC -1- INDUCTION: By phenobarbital.
 CC -1- PTM: N-GLYCOSYLATED (PROBABLY).
 CC -1- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF THE LIVER ISOZYME.
 CC -1- THE KIDNEY ISOFORMS DIFFERS IN 12 POSITIONS.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).
 CC -----
 DR EMBL: U06273; AAA83404.1; -;
 DR EMBL: U06274; AAA83405.1; -;
 DR PIR: S68200; S68200.
 DR InterPro: IPR002213; UDP-glucosyltrans.
 DR Pfam: PF00201; UDPGT_1.
 DR ProSite: PS00375; UDPGT_1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KW Multigene family; Microsome.
 FT SIGNAL 1 23
 FT CHAIN 24 530 UDP-GLUCURONOSYLTRANSFERASE 2B12.
 FT TRANSMEM POTENTIAL.
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 2 2 S -> P (IN KIDNEY).
 FT VARIANT 61 61 F -> S (IN KIDNEY).
 FT VARIANT 71 71 D -> H (IN KIDNEY).
 FT VARIANT 95 95 N -> S (IN KIDNEY).
 FT VARIANT 183 183 Q -> K (IN KIDNEY).
 FT VARIANT 346 346 P -> T (IN KIDNEY).

FT VARIANT 398 398 A -> G (IN KIDNEY).
 FT VARIANT 414 415 VE -> AT (IN KIDNEY).
 FT VARIANT 433 433 V -> D (IN KIDNEY).
 FT VARIANT 475 475 K -> L (IN KIDNEY).
 FT VARIANT 488 488 Q -> L (IN KIDNEY).
 SQ SEQUENCE 530 AA; 6106 MW; D49313CE3B6D5BED CRC64;
 Query Match 69.0%; Score 1930.5; DB 1; Length 530;
 Best Local Similarity 66.8%; Pred. No. 1,36-138;
 Matches 354; Conservative 73; Mismatches 102; Indels 1; Gaps 1;
 QY 1 MSMTATALLLIQLSCYFSSGCGKVLVWPFEBSHNNIKITLDELVQRHEVTVLASSA 60
 DB 1 MSGKISALLLIQISFCFKSGCKGLVWPWEYSHNNIKITLDELVQRHEVTLPSSA 60
 QY 61 SISFDPSPSTLKEVYPVSLTKTEFSDIIKQVKKRA-ELPKDTPMSYSQVGEIMWT 119
 DB 61 FVLDPKETSCLKVTPTPTSSSHDENPFTFRFVNVVTELPRTCLSYFLYLDOTIDY 120
 QY 120 NDILRRCCKDIVSKMLKLOESRPDVLAADVPEGELALDLKIPVYSLRPSGYA 179
 DB 121 SDYCLTYCKEAVSNKQMTKLQESKFDVPSDALPGGELALDLKIPVYSLRPSGYT 180
 QY 180 TEKHSGLLEPPSYVPVMSSELSQMTPIERKMYLYLFEFVQIPLMKKMDQFSEV 239
 DB 181 IEQYIGVLPSPSYVPMIFSGLAGQMTFIERVHNMICLYEDFWQFREKKMDPFYSKT 240
 QY 240 LGRPTLTSETAKADILIRVWDFOPPHLLPVVEFVGILCKPAKPLPKMEPEVOSS 299
 DB 241 LGRPTTLAETNGKEMWLIRVWDFOPPHLLPVVEFVGILCKPAKPLPKDIEDFVOSS 300
 QY 300 GENGVVFSLSGMYSNTESEERANVIAALAKIPQKVLMPFGNKPDTLGLNTRLYKNIPQ 359
 DB 301 GEHGVVFSLSGMYSNTESEERANVIAALAKIPQKVLMPFGNKPDTLGLNTRLYKNIPQ 360
 QY 360 NDILGHPKTKAFTHGGNGNGIYEAIYGVPMVGVPIFGDQLDNIAHAKAAGAEINFT 419
 DB 361 NDILGHPKTKAFTHGGNGNGIYEAIYGVPMVGVPIFGDQLDNIAHAKAAGAEINFT 420
 QY 420 MTSDDLRALETVTDDSSYKNAARLSRIHHDQVKKPLDRVPMIEFVMRKGAAGHLSA 479
 DB 421 MTKSDLLNALEEVINDPYNKNAWMLSTIHDQPTKPLDRVPMIEFVMRKGAAGHLSL 480
 QY 480 AHDLTWFQHSIDVIGFLTCVATAIFLTKCFLFSCQKFNKTRIEKRE 529
 DB 481 GNLPMWQYHSLDVIGFLSCVAATVVALKCFVYVRFVKEKKTKNE 530

Search completed: December 5, 2003, 09:48:50
 Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:47:25 ; Search time 21 Seconds
(without alignments)
1065.831 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSKMWTSAALLIOLSCYFSS.....KCFLLFSCQKFNKTRKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents AA:
2: /cgn2_6/ptodata/1/aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/aa/PTUTS.COMB.pep.*
7: /cgn2_6/ptodata/1/aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2602.5	93.0	528	4	US-09-356-806-8
2	2322	83.0	524	4	US-09-356-806-8
3	2169.5	77.5	530	3	US-09-180-852-2
4	2158.5	77.1	530	4	US-09-356-806-113
5	1874.5	67.0	454	4	US-09-813-918-2
6	1243	44.4	288	4	US-09-813-918-3
7	1196.5	42.8	531	5	PCT-US92-00282-6
8	1130.5	40.4	533	5	PCT-US92-00282-3
9	1126	40.2	534	5	PCT-US92-00282-4
10	1064	38.0	531	5	PCT-US92-00282-7
11	1045.5	37.4	529	5	PCT-US92-00282-7
12	833	29.8	197	4	US-09-813-918-4
13	765.5	27.4	245	4	US-09-305-856B-18
14	403	14.4	515	3	US-08-942-012B-32
15	393	14.0	129	4	US-09-370-838-36
16	379	13.5	488	3	US-08-942-012B-29
17	379	13.5	488	3	US-08-942-012B-30
18	374	13.4	460	3	US-08-942-012B-33
19	372	13.3	288	5	US-09-305-856B-2
20	372	13.3	288	5	PCT-US92-00282-19
21	362.5	13.0	289	4	US-09-305-856B-8
22	360.5	12.9	289	4	US-09-305-856B-6
23	360.5	12.9	289	5	PCT-US92-00282-13
24	358.5	12.8	289	4	US-09-305-856B-4
25	358.5	12.8	289	4	PCT-US92-00282-15
26	354.5	12.7	289	5	PCT-US92-00282-11
27	346	12.4	310	4	US-09-305-856B-14

28	338	12.1	515	3	US-08-942-012B-24	Sequence 24, Appl
29	321.5	11.5	317	4	US-09-305-856B-12	Sequence 12, Appl
30	316	11.3	287	4	US-09-305-856B-10	Sequence 10, Appl
31	306.5	11.0	286	5	PCT-US92-00282-9	Sequence 9, Appl
32	297.5	10.6	98	5	PCT-US92-00282-26	Sequence 26, Appl
33	296.5	10.6	253	4	US-09-305-856B-16	Sequence 16, Appl
34	294.5	10.5	506	3	US-08-942-012B-26	Sequence 26, Appl
35	287	10.3	506	3	US-08-942-012B-25	Sequence 25, Appl
36	287	10.3	506	6	5180581-2	Patent No. 5180581
37	271	9.7	58	2	US-08-466-583-9	Sequence 9, Appl
38	271	9.7	58	4	US-08-265-427-9	Sequence 9, Appl
39	271	9.7	58	5	PCT-US95-07820-9	Sequence 9, Appl
40	261	9.3	493	3	US-08-942-012B-28	Sequence 28, Appl
41	257	9.2	52	2	US-08-466-583-7	Sequence 7, Appl
42	257	9.2	52	4	US-08-265-427-7	Sequence 7, Appl
43	257	9.2	52	5	PCT-US95-07820-7	Sequence 7, Appl
44	250.5	9.0	489	3	US-08-942-012B-31	Sequence 31, Appl
45	241.5	8.6	491	3	US-08-942-012B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1		US-09-356-806-8	Sequence-A: Application US/09356806
		Patent No. 6586175	
GENERAL INFORMATION:			
APPLICANT: Penny, Laura			
APPLICANT: Galvin, Margaret			
APPLICANT: Miller, Andrew			
APPLICANT: Reidy, Michael			
TITLE OF INVENTION: Genotyping Human			
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and			
FILE REFERENCE: 2B15 (UGT2B15) Genes			
CURRENT APPLICATION NUMBER: US-22PRV2			
CURRENT FILING DATE: 1999-07-20			
NUMBER OF SEQ. ID NOS: 164			
SOFTWARE: FASTSEQ for Windows Version 3.0			
SEQ ID NO: 8			
LENGTH: 528			
TYPE: PRT			
ORGANISM: H. sapiens			
US-09-356-806-8			
Query Match	93.0%	Score 2602.5	DB 4; Length 528;
Best Local Similarity	93.0%	Pred. No. 4.7e-263	
Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;			
QY	1	MSKMWTSAALLIOLSCYFSSGCKVLVWPTFSHMNITITIDELVORGHETVVLASSA	60
DB	1	MSKMWTSAALLIOLSCYFSSGCKVLVWPTFSHMNITITIDELVORGHETVVLASSA	60
QY	61	SISDPSPSTLKEFVYVSLTTEFEDIKOLVKRAELPKDTFMSYFSOVQOIMTFTN	120
DB	61	SISDPSPSTLKEFVYVSLTTEFEDIKOLVKRAELPKDTFMSYFSOVQOIMTFTN	120
QY	121	DIRKFKCDIVSNKKMLKQIESRFVDVLA DAVPFEGELAEILKIDPVYSLRSPGYAI	180
DB	121	DIRKFKCDIVSNKKMLKQIESRFVDVLA DAVPFEGELAEILKIDPVYSLRSPGYAI	180
QY	181	EKISGGLFPSPYVPMVMSLSDMTIERVKMITYLYEFPMQITDMKKMQDFYSEVL	240
DB	181	EKISGGLFPSPYVPMVMSLSDMTIERVKMITYLYEFPMQITDMKKMQDFYSEVL	240
QY	241	GRPTTSETMAKADIMILIRYVDFOPHPLLPVNEFVGLHCKRPKPLPKMEEFVSSG	300
DB	241	GRPTTSETMAKADIMILIRYVDFOPHPLLPVNEFVGLHCKRPKPLPKMEEFVSSG	300
QY	301	ENGWVPSLSQWVSNTESEERANYIASALAKIPQVLRFPDGNKPDITGLNTRLYKWI	360
DB	301	ENGWVPSLSQWVSNTESEERANYIASALAKIPQVLRFPDGNKPDITGLNTRLYKWI	360

Query Match 83.0%; Score 2322; DB 4; Length 524;
Best Local Similarity 81.5%; Pred. No. 9.7e-234;
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0;

361 DLGHPKTAFTTHGGMNGIYEAIYHGVPMVGPPIFGDLDNIAMHKAAGAIVEINFKT 420
361 DLGHPKTAFTTHGGMNGIYEAIYHGVPMVGPPIFGDLDNIAMHKAAGAIVEINFKT 420
421 TSEDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVNRHKGAKILRSAA 480
421 SSTDLLNKLKTVINDPLYKENAMKLSRIHHDPVKPLDRAVFWIEFVNRHKGAKILRSAA 480
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 529
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 529

RESULT 2
US-09-356-806-40
Sequence 40, Application US/09356806
Patent No. 6586175
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22BRV2
CURRENT APPLICATION NUMBER: US/09/356,806
CURRENT FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 40
LENGTH: 524
TYPE: PRT
ORGANISM: H. sapiens
US-09-356-806-40

Query Match 83.0%; Score 2322; DB 4; Length 524;
Best Local Similarity 81.5%; Pred. No. 9.7e-234;
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0;

361 DLGHPKTAFTTHGGMNGIYEAIYHGVPMVGPPIFGDLDNIAMHKAAGAIVEINFKT 420
361 DLGHPKTAFTTHGGMNGIYEAIYHGVPMVGPPIFGDLDNIAMHKAAGAIVEINFKT 420
421 TSEDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVNRHKGAKILRSAA 480
421 SSTDLLNKLKTVINDPLYKENAMKLSRIHHDPVKPLDRAVFWIEFVNRHKGAKILRSAA 480
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 529
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 529

Query Match 77.5%; Score 2169.5; DB 3; Length 530;
Best Local Similarity 74.7%; Pred. No. 8.6e-218;
Matches 396; Conservative 60; Mismatches 73; Indels 1; Gaps 1;

481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 524
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 524

RESULT 3
US-09-180-852-2
Sequence 2, Application US/09180852
Patent No. 6287834
GENERAL INFORMATION:
APPLICANT: BELANGER, Alain
APPLICANT: HUM, Dean W.
APPLICANT: BEAULIEU, Martin
APPLICANT: LEVESQUE, Eric
TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
TITLE OF INVENTION: DIHOSPHO-GLUCURONOSYLTRANSFERASE
FILE REFERENCE: 1259-449
CURRENT APPLICATION NUMBER: US/09/180,852
CURRENT FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: PCT/CA97/00328
EARLIER FILING DATE: 1997-05-16
EARLIER APPLICATION NUMBER: US 08/649,319
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 530
TYPE: PRT
ORGANISM: Homo sapiens
US-09-180-852-2

Query Match 77.5%; Score 2169.5; DB 3; Length 530;
Best Local Similarity 74.7%; Pred. No. 8.6e-218;
Matches 396; Conservative 60; Mismatches 73; Indels 1; Gaps 1;

481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 524
481 HDLTFQYHSIDVIGFLLTCVATAFLFTKCFLESCQKFNKTRKEKE 524

RESULT 4


```

US-09-356-806-113
; Sequence 113, Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; FILE REFERENCE: SEQ ID NO 2
; CURRENT APPLICATION NUMBER: US/09/356,806
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 530
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-356-806-113

Query Match      77.1%; Score 2158.5; DB 4; Length 530;
Best Local Similarity 74.5%; Pred. No. 1.2e-216;
Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQISCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSKMTSVFLIIQISCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60

QY 61 SISFDPSPETLKEEYVPSLTKEPEDIIKOLVKRMA-ELPKDTFMSYFQVOEIMWTF 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 STLVNAKSSAIKIEVYPTSLTKNDLSDLIKIDRIYGSKTFMSYFQVOELCHEY 120

QY 120 NDILRKECKDIVSNKKMLKQLQESRFDVVLADAVFPCELLAEELKIPVYSLRFPQYA 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 YDVSNKLCKDAVLNKKLMKKLQESKFDVILADALNPGCELLAEELNIPVYSLRFSVGYT 180

QY 180 IEKSGGLLPSPVYVPMVMSLSQMTFIERVKMIVLYLVEFQFQIFDMKKMQFYSEV 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 FEKGGGGLFPSPVYVPMVMSLSQMTFIERVKMIMHLYFDFWFOYDLKKMQFYSEV 240

QY 240 LGRRPTLSEMAKADIMLRNVMDFOPPHPLPNVEFVGGLHCKPAKPLPKEMEFPVQSS 299
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 LGRRPTLSEMAKADIMLRNVMDFOPPHPLPNVEFVGGLHCKPAKPLPKEMEFPVQSS 300

QY 300 GENGCVVFSIGSMVSNTESEERANYIASALAKIPQKVLWRPDKNPKDTLGLNTRLYKWIPO 359
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 GENGIVVFSIGSMVSNTESEERANYIASALAKIPQKVLWRPDKNPKDTLGLNTRLYKWIPO 360

QY 360 NDILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDQDNTAHMKAKAAVEINFKT 419
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 NDILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDQDNTAHMKAKAAVEINFKT 420

QY 420 MTSDDLRLATVITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 479
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 MTSDDLRLATVITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 480

QY 480 AHDLTWFQHSIDVIGFLITCVATAIFLTKCPLFSCCKPKNTRKIERRE 529
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 AHDLTWFQHSIDVIGFLITCVATAIFLTKCPLFSCCKPKNTRKIERRE 530

RESULT 5
US-09-813-918-2
; Sequence 2, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF

```

```

; FILE REFERENCE: CL0001175
; CURRENT APPLICATION NUMBER: US/09/813,918
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Human
US-09-813-918-2

Query Match      67.0%; Score 1874.5; DB 4; Length 454;
Best Local Similarity 67.9%; Pred. No. 4.6e-187;
Matches 359; Conservative 35; Mismatches 60; Indels 75; Gaps 1;

QY 1 MSKMTSALLLIQISCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MTKMTSVLLIHLSCYFSSGCGKVLVMAEYSHMMNMTITIKELVORGHETVLAASA 60

QY 61 SISFDPSPETLKEEYVPSLTKEPEDIIKOLVKRMAELPKDTFMSYFQVOEIMWTFN 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SILFDPNDASTLKEEYVPTSLTKPEENITMOQVKRMSDKRKSFWLYFQBOEILMELY 120

QY 121 DILRKECKDIVSNKKMLKQLQESRFDVVLADAVFPCELLAEELKIPVYSLRFPQYAI 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 DIFRNFCKDIVSNKKMLKQLQELRPDIVADAVFPCELLAEELNIPVYSLRFSVGYT 166

QY 181 EKSGGLLPSPVYVPMVMSLSQMTFIERVKMIVLYLVEFQFQIFDMKKMQFYSEV 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 ----- 166

QY 241 GRPTLSEMAKADIMLRNVMDFOPPHPLPNVEFVGGLHCKPAKPLPKEMEFPVQSSG 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 -RPTTLSEMAKADIMLRNVMDFOPPHPLPNVEFVGGLHCKPAKPLPKEMEFPVQSSG 225

QY 301 ENGCVVFSIGSMVSNTESEERANYIASALAKIPQKVLWRPDKNPKDTLGLNTRLYKWIPO 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 226 ENGCVVFSIGSMVSNTESEERANYIASALAKIPQKVLWRPDKNPKDTLGLNTRLYKWIPO 285

QY 361 DILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDQDNTAHMKAKAAVEINFKT 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 286 DILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDQDNTAHMKAKAAVEINFKT 345

QY 421 TSEDLLRLATVITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 346 TSEDLLRLATVITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 405

QY 481 HDLTWFQHSIDVIGFLITCVATAIFLTKCPLFSCCKPKNTRKIERRE 529
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 406 HDLTWFQHSIDVIGFLITCVATAIFLTKCPLFSCCKPKNTRKIERRE 454

RESULT 6
US-09-813-918-3
; Sequence 3, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL0001175
; CURRENT APPLICATION NUMBER: US/09/813,918
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Human
US-09-813-918-3

```

Query Match	44.4%	Score 1243;	DB 4;	Length 288;
Best Local Similarity	80.2%;	Pred. NO. 2.3e-121;		
Matches 231;	Conservative 21;	Mismatches 36;	Indels 0;	Gaps 0;

Qy	242	RPTLSEMAKADIMLIRBYMDFOPHPLPLVNEVFGJHCKRAKPLPREMEFEVSSGE	301
Db	1	RPTLFEHMGKADIMLRNMSGFPORPHPLPVNDFVGJHCKRAKPLPREMEFEVSSGE	60
Qy	302	NGVVVFSLSGWSVNTSEERANYIASALAKIPOKVLVRFPGNKPDTLGLNTRLYKWI	POND 361
Db	61	NGVVVFSLSGVSNNMTAERANYIATLAKIPOKVLVRFPGNKPDAJGLNTRLYKWI	POND 120
Qy	362	LLGHFKTAFTIHGGANGIYEAIYHGVPMWGPVIFEGDOLNIAHMAKGAAYINFERTMT	421
Db	121	LLGHFKTAFTIHGGANGIYEAIYHGI PMWGPVIFPDQDNI AHMAKGAAYLDNRTMS	180
Qy	422	SEDLRLALRTVITDSSYKENAMRLSRHHDDQPKPLDRAVFWIEFVWRHKGAYLSAAH	481
Db	181	STDLNNAKTVINDEPLYKENIMKLSRICHDDQPKPLDRAVFWIEFVMPHGAKGLRVAAH	240
Qy	482	DLTWFOHNSDIVIGFLTLVATAIPLFTQCPLFSQCFKPAKTRIEKRE	529
Db	241	DLTWFOHNSDIVIGFLTLCVATVIFITITFCLEFCFMPKFAKKGKGRD	288

RESULT 7

PCT-US92-00282-6
; Sequence 6, Application PC/TUS9200282

APPLICANT: OWENS, IDA S
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-6

[illegible]

```

Db      / LR6SGJLLLLCALPMA66--GKVLVFM6SGHMLSMNDVVR6LHAHQAVLAP6TV 64
Oy      63 SFPDPSBPSTLKEEVYVFSVLKTEEFDDITKOLYKRNALPKDTFMSYSQVOELMTTNDI 1222
Db      65 HMKGBEDFTLQIOTAF6--YTK6EYORELIGNAKK6FE-POH6VVT6FEETMASIK-KF6D 1200
Oy      123 LRKFC6DVI6NKKYKMLKQ6ESRFDVLADAVEPFE6LAE6LTKIPFY6SLR6SGVI6K 1828
Db      121 YANS6CALL6HNTLITLO6LSS6FDVVLDP6VPC6ALLAKTLOIP6VFLR6SV6-6CIDY 1797
Oy      183 HSG6LL6PP6SY6PV6VMS6LSDMT6PI6ERV6KNI6YVL6FE6MFQID6KK6MDQ6Y6SVL6R 2422
Db      180 E6TQ6CP6ES6SYIP6LNTL6MLSDMT6L6V6K6MLY6LTK6Y6ICH6-SITP6Y6SL6S6ELL6R 2388
Oy      243 PTL6S6M6K6DI6WL6RN6WD6Q6PH6LPI6NVE6F6G6H6CK6AP6RPL6K6ME6EP6Q6SG6N 3022
Db      239 E6MS6LVE6YLS6AS6YML6RR6D6P6VD6P6R6P6IM6PN6V6FG6INC6Y6IK6P6L6S6E6F6AY6N6SG6H 2989
Oy      303 G6VV6F6SL6GS6V6N6T6SE6R6N6Y6AS6AL6KI6P6K6VL6R6PD6N6K6PDTL6G6N6TR6LY6KI6P6NDL 3622
Db      299 GI6VF6SL6GS6V6S6EIP6E6KK6ME6IA6BL6GRIP6LL6R6Y6T6G6T6R6P6ML6AK6NTLITL6VK6LP6NDL 3568
Oy      363 LG6R6PK6AFIT6TG6M6GI6Y6E6AL6YH6GV6PW6GV6I6ED6Q6D6N6I6AH6K6A6G6A6V6E6N6FK6MTS 4222
Db      359 LG6R6PK6AFIT6H6SG6I6Y6E6GIC6N6GV6PW6M6V6L6ED6Q6D6N6K6R6E6T6G6AG6Y6TL6V6L6E6TA 4188
Oy      423 ED6LR6AL6R6L6T6D6S6Y6K6EN6AM6LS6RI6IH6D6Q6AP6K6P6DR6AV6N6I6E6F6V6R6NH6K6A6H6LS6A6HD 4882
Db      419 DDL6EN6L6K6Y6I6N6K6S6Y6K6EN6I6M6R6LS6L6H6DR6PI6E6L6DL6AV6F6V6E6Y6V6M6NH6K6A6P6H6R6A6HD 4788
Oy      483 L6TF6O6H6SID6V6IG6FL6T6CAT6AIF6L6FT6C6FL6SC6OK-P6N6K6R6K6I6E6K 527
Db      479 L6TY6O6Y6H6SD6V6IG6FL6A6I6VL6V6F6I6V6Y6S6CA6G6CK6CG6G6K6R6Y6K 524

```

RESULT 8

PCT-US92-00282-3
: Sequence 3, Application PC/TUS9200282

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-3

Query Match 40.4%; Score 1130.5; DB 5; Length 533;
Best Local Similarity 44.7%; Pred. No. 3,56-109;
Matches 232; Conservative 89; Mismatches 163; Indels 35; Gaps 7;

24 GKVLWPTFESHMNNIKITIDELVORGHEVTLASSASISFDPSPSTLKEVVP----- 78
28 GKILILVDSGSHMISGALQLOOQRGHEIVLAPDASLYRDAFYTLK--TYVPPQR 85
79 -----VSLTKTEFED--IKQLVKRMALPKDTFWSYFQVOEIMTFNDILRKCKD 129
86 EDVESFVSLGHNVFENDSFQRIYIKYKIKD-----SAMLSCSH 129
130 IVSNKKIMKKLOESRFDVVLADAVPPGELLAEILKIPFYSLRPSQYALKEHSGLLF 189
130 LHNKEELMASLAESSFVMLTDPPLPCSPVAQYLSPTVFFHLALP-CSLEFATQCPN 188
190 PPSVPPVMSLSQMTFIERVKMIVLYFEFVQIFDMKKWQFSEVLGRTTLSET 249
189 PFSVPPPLSHSDHMTFLQVKMLIAFSQNFCDVW-YSPYATLASFLQREVTQDL 247
250 MAKADILIRYWDFOFPHPLPVEFVGGLHCKPAKPLPKMEEFVQSSGENGVVFSL 309
248 LSSASVLFPSDFKQVPRIPMPMVGVGINCLHONPLSOEFAYINASEGHIYVPSL 307
310 GSVVNTSEERANVIALAKIPOKVLMPDGNKPDTLGLNTRLYKMI PONDLLGHPKTK 369
308 GSWSEIPEKKAMALADALGKNPQTVLMRYGTRPSNLANTLILVKMLPQNDLLGHPMTR 367
370 APTHGMMNGIYEAIHYGVPMVGPPIFGDQDNLIAHMKAKGALEINFKMTSDDLRL 429
368 APTHAASHGYEESI CNGVPMVMPPLFGDDQDNKRMETKAGATLVLENTSSDLNAL 427
430 RTVITDSSYKENARLSRIHHDQVVKPLDRAVFWIEFVMRHKGAKHLSAHDLTWFOHY 489
428 KAVINDSYKENIMRLSLHKDRVEPLDLAVFVEFMRHKGAPHLRAHDLTWQYH 487
490 SIDVIGFLTCVATAIFLFTKCFLPSQCK-FNKTRKIEK 527
488 SLDIVIGFLAVLTVAFITFKCAVGRKCLGKGRVVK 526

RESULT 9
PCT-US92-00282-4
Sequence 4, Application PC/TUS9200282
GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-4

Query Match 40.2%; Score 1126; DB 5; Length 534;
Best Local Similarity 43.9%; Pred. No. 1e-108;
Matches 234; Conservative 93; Mismatches 176; Indels 30; Gaps 7;

6 TSALLILQLSCYSSGCGKVLWPTFESHMNNIKITIDELVORGHEVTLASSASISFD 65
14 TGLLLLSVPMALSS---GKVLVPTDGSPLMSKRELRHARGHQAVALTEPVNMHIK 70
66 PNPSTLKFEVYVSLTKTEFEDI-----IKQLVKRMALPKDTFWSYFQVOEI 115
71 EEKFFTL--TAVAVPMTQKFDRTVLGYTQGFETELHKLK-----YSRMAI 116
116 MTFPNDILRKFCQDIYSNKKLMKKLOESRFDVVLADAVFPFGLLAEILKIPFYSLRFS 175
117 MNVSLALHRCVCELLNEALIRHLNATSPDVLTDPVNLGAVLAKYLSIPAVFWRYI 176
176 PGYAIKHSGLLFPPSYVVMVMSLSQMTFIERVKMIVLYFEFVQIFDMKKWQF 235
177 P-CDLPKQTCQCNPSYITKLLTNSDHTFLQRYKMMIYPLALSTICHTFS-APYASL 234
236 YSEVLGRTTLSETMAKADILIRYWDFOFPHPLPVEFVGGLHCKPAKPLPKMEEF 295
235 ASEIFQGEVAVDVLVSASVWLPRGDPVMDYPRIPMNVFISGINCANGKPSQEPY 294
296 VOSSGENGVVPSLGSVNTSEERANVIALAKIPOKVLMPDGNKPDTLGLNTRLYK 355
295 INASGEHIVFSLGSVSEIPEKKAMALADALGKIPQVLMRYGTRPSNLANTLILVK 354
356 WIPONDLLGHPKTKAFTTHGMMNGIYEAIHYGVPMVGPPIFGDQDNLIAHMKAKGALEI 415
355 WIPONDLLGHPMTRAITTHAGSHGYEESICNGVPMVMPPLFGDDQDNKRMETKAGAVTL 414
416 NFKMTSDDLRLRLRTVITDSSYKENARLSRIHHDQVVKPLDRAVFWIEFVMRHKGAKH 475
415 NVLEMTSEDLNALKAVINDSYKENIMRLSLHKDRVEPLDLAVFVEFMRHKGAPHL 474
476 LRSAHDLTWFOHYSIDVIGFLTCVATAIFLFTKCFLPSQCK-FNKTRKIEK 527
475 LRPAHDLTWQYHSDIVIGFLAVLTVAFITFKCAVGRKCLGKGRVVK 527

RESULT 10
PCT-US92-00282-5
Sequence 5, Application PC/TUS9200282
GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
PCT-US92-00282-5

Query Match      38.0%; Score 1064; DB 5; Length 531;
Best Local Similarity 43.4%; Pred. No. 3.1e-102;
Matches 221; Conservative 90; Mismatches 180; Indels 18; Gaps 8;

25 KVLVWPTFESHWNIKITLDELVORGHEVTVLASSASISFDPNSPSTLKFEYVSVLTKT 84
28 KLLVVPQSGSHWLSKMDIVEVLSDRGHEIVVVPVNNLLKRYKYTRK--IYVPVYDOE 85
85 EFEDIIKOLVYKMALEPKDTFWS-YFSQVQEIIMTFNDILRKF-----CKDIVSNKKMK 139
86 E-----LKKRYGSGFNHFAERSFLTAPOTEYNNMIVGLYFNCQSLQDDTLNF 138
140 LQESRFDVLADAVFPFGLAELKIPVYSLRSPGVALEKHSGLLPSPSYVPVMS 199
139 FKESRFDALFTDPALPCGVIIAELVGLSVYLFRRFP--CSLEHTSRSPDPVSYIPRCYT 197
200 ELSDQMTFIERVKMNIYVLYEEFWFOIFDMKKMDQFSEVILGRPTLTSETMAKADIWLR 259
198 KFSDMHTESQRAVNLVNLBEPYLCU-FSKYEKLAGAVLKRVDII-TLSEVSVMLDLR 255
260 NYWMDQFPHPLBNPEFVGLHCKRAKPLPKEMEBFVQSSGNGVNVVSLGSMVSNTEE 319
256 YDFVLEYRPPVMPNVPFGLGNCCKRKQDSQEFVAVINASGEHGLVSVLSGWSVEIPEK 315
320 RANVASALAKIPQKVLWRFQGNKPDTLGLNTRLYKMI-PONDLLGHPKTKAFITTHGMNG 379
316 KAMALADALGNPQVULNRYGTRPSNLANTILVKMLPONDLLGHPMTTRAITTHAGSHG 375
380 IYEALYHGVNVPVPIFGDDLDNIAHKAAGAIVEINFKMTSEDLRLALRTVITDSSYK 439
376 VYESICNGVPMVMDLFGDDMDNAKMETKAGAVTLNVLMTSEDLNALKAVINDKSYK 435
440 ENAMLSRIHDPKPLDRAVPMIEFVWRHGAHLASAADLTWFOHYSIDVIGPLLT 499
436 ENIMLSLSLHKDRPVEPLDLAVFWVEFVWRHGAHLRAADLTWYOHSLDIVIGFLLA 495
500 CVATAIPLFTKCFLFSCOK-FNKTRKIEK 527
496 VVLTAIPLFTKCFPGYKCGKCGKGRVKK 524

```

```

NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
PCT-US92-00282-7

Query Match      37.4%; Score 1045.5; DB 5; Length 529;
Best Local Similarity 41.1%; Pred. No. 2.7e-100;
Matches 211; Conservative 101; Mismatches 174; Indels 27; Gaps 7;

25 KVLVWPTFESHWNIKITLDELVORGHEVTVLASSASISFDPNSPSTLKFEYVSVLTKT 84
27 KLLVVPQSGSHWLSKMDIVEVLSDRGHEIVVVPVNNLLKRYKRSKPPVYNL--- 83
85 EFEDIIKOLVYKMALEPKDTFWSYFSQVQEIIMTFNDILRKF-----CKDIVSNKK 135
84 -----ELRRRYNSFGN---NHFAASSPLMAPLREYNNMIVIDMCFSCQSLKQSA 133
136 LMKLQESRFPVLADAVFPFGLAELKIPVYSLRSPGVALEKHSGLLPSPSYVP 195
134 TLFSLRENOFPALFTDPAMPQGVIIAELYKLSIYLFRRGP--CSLEHIGQSPSPSVYP 191
196 VVMSLSDQMTFIERVKMNIYVLYEEFWFOIFDMKKMDQFSEVILGRPTLTSETMAKADI 255
192 RPYTKFSDMHTFQRLAFLNINLENTLYHCL-YSKTEILASDLKQDVS-LPALHONS 249
256 WLIRNYMDFQPHPLBNPEFVGLHCKRAKPLPKEMEBFVQSSGNGVNVVSLGSMVSN 315
250 WLLAYDFEYRPPVMPNVPFGLGNCCKKKNLQSEFVAVYNASGEHGLVSVLSGWSVE 309
316 TSEERANVASALAKIPQKVLWRFQGNKPDTLGLNTRLYKMI-PONDLLGHPKTKAFITTHG 375
310 IPEKKAMETIABALGRIPQTLIMRYGTRPSNLAQNTILVKMLPONDLLGHPKAKAFITHS 369
376 GNGIYELIYHGVNVPVPIFGDDLDNIAHKAAGAIVEINFKMTSEDLRLALRTVITD 435
370 GSHGIYEGICNGVPMVMDLFGDDMDNAKMETKAGAVTLNVLMTADLADNALKTYVANN 429
436 SSYKENAMLSRIHDPKPLDRAVPMIEFVWRHGAHLASAADLTWFOHYSIDVIG 495
430 KSYKENIMRSLSLHKDRIEPLDLAVFWVEFVWRHGAHLRAADLTWYOHSLDIVIG 489
496 FLITCVATAIPLFTKCFLFSCOK-FNKTRKIEK 527
490 FLTAIPLTVFVIYKSCAYGCRKCFGKGRVKK 522

```

```
RESULT 12
US-09-813-918-4
; Sequence 4, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; FILE REFERENCE: CL001175
; CURRENT APPLICATION NUMBER: US/09/813,918
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Human
US-09-813-918-4

Query Match          29.8%; Score 833; DB 4; Length 197;
Best Local Similarity 78.2%; Pred. No. 8.8e-79;
Matches 154; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSMKWTSAALLIQLSCYSSSGCGKVLWPTFESHMNIKITLDELVQRGHEVTLASSA 60
DB 1 MTLKMTSVLLIHLISCYFSSGCGKVLWMAEYSHMMNMKITLDELVQRGHEVTLASSA 60
QY 61 SISDPNPSPTLKEPVVSLTKTEFEDIIKQLVKRWAELEPKDTFMSYFSQVOEIMTPTN 120
DB 61 SILFDPNDASTLKEVPTSLTKTEFENIIQOVKMSDIKDSFWLYFSQOELLWELY 120
QY 121 DILKRFCKDIYSNKKMLKQLQESRPDVLADAVFPFGLLAELIKIPVYSLRFSPGYAI 180
DB 121 DIFNFCDDVYNSKKMKKQLQESRPDIYFADAVFPFGLLAELINIRVYSLRFTPGYTI 180
QY 181 EKHSGLLFPSPYV 197
DB 181 ERHSGGLIFPPSYIPV 197

RESULT 13
US-09-305-856B-18
; Sequence 18, Application US/09305856B
; Patent No. 6479236
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; TITLE OF INVENTION: Genotyping the Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 1 (UGT1) Gene
; FILE REFERENCE: 4389-7 (formerly SEQ-17CIP)
; CURRENT APPLICATION NUMBER: US/09/305,856B
; CURRENT FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 60/084,807
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-305-856B-18

Query Match          27.4%; Score 765.5; DB 4; Length 245;
Best Local Similarity 60.1%; Pred. No. 1.4e-71;
Matches 143; Conservative 35; Mismatches 59; Indels 1; Gaps 1;

QY 291 EMEFVQSSGNGVYVSLGSMVNTSEERANVIASALAKIPQVLMRFDGNKPDPLGLN 350
DB 1 EFAFYINASGBHGIIVFSLGSMVSEIPEKKAMIALDALGKIPTVLMRYTGTSPRSNLANN 60
```

```
QY 351 TRLVKWIPONDLLGHPKTKAFITGGNAGIYEAIYHGVPMVGVPIFGDOLDNIAHMAKAG 410
DB 61 TILVKMLPONDLLGHPKTRAFITHAGSHGYEISCNQVPMVMPLEFDQDNKAKMETKG 120
QY 411 AAVEINKTISDILLALRTVITDSSYKENAMRLSIHHDDPKPLDRAVMEIYVWRH 470
DB 121 AGTVLNLVEMTSBDELEVALKAVINDSKYKENIMRLSLHHKDRPEPLDLAVFWEFVWRH 180
QY 471 KGAKHLRSAAHDLTPQYHSIDYIGPLTCAVAIAIPLFTGCFPSOCK-FNKRIRKIK 527
DB 181 KGAPHLRPAHDLTWYQYHSLDVIIGFLAVVLTVAFTFFCCAYGYRCKLGKGRVAK 238

RESULT 14
US-08-942-012B-32
; Sequence 32, Application US/08942012B
; Patent No. 6235278
; GENERAL INFORMATION:
; APPLICANT: Lu, Albert
; APPLICANT: Miller, Lois K.
; APPLICANT: Dietze, Peter
; APPLICANT: Black, Bruce
; TITLE OF INVENTION: Biological Insect Control Agents Expressing
; TITLE OF INVENTION: Insect-Specific Toxin Genes, Methods and Compositions
; FILE REFERENCE: 28-96a
; CURRENT APPLICATION NUMBER: US/08/942,012B
; CURRENT FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: 08/729,606
; PRIOR FILING DATE: 2000-10-01
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Spodoptera littoralis nuclear polyhedrosis virus
US-08-942-012B-32

Query Match          14.4%; Score 403; DB 3; Length 515;
Best Local Similarity 26.7%; Pred. No. 3.7e-33;
Matches 141; Conservative 102; Mismatches 227; Indels 58; Gaps 20;

QY 9 LLLIQLSCYSSSGCGKVLWPT-FESHMNIKITLDELVQRGHEVTLASSASISDPN 67
DB 5 ILVVSILVLRNSAAVRLCMFTPTPSYSHQVFPDYVVALLRHSHLVIT--SKPIHHNH 62
QY 68 SPSTLKEVYVSLTKTEFEDIIKQLVKRWAELEPKDTFMSYFSQVOEIMTPTDILR--- 124
DB 63 GHRHHRHE---NLTEIDVGSVTNNPFKR---LLQD---SKVSRKGIYSDSSTVTRVNY 112
QY 125 -----KFCDDIYSNKKMLKQLQESRPDVLADAVFPFGLLAELIKIPVYSLRFS 175
DB 113 LGLARMISAQFEHQV--KRLRSNQ--TPDVIVIEAFVSPILSYFFDTPV--IQIS 166
QY 176 PGYAIKQ---SGGLLFPSPYVPMVMSSELDQMTFIERVAGMI--VYLVEFPMQLPD-- 228
DB 167 SGHGTAEFTMGAAVAPVYVPMVMDRFRKGLSWQCTQVTEIRLILYFESQLDADQS 226
QY 229 --KKWDQFYSEVILGPTTLETMAKADILIRNYMDFOPPHPLPVEVFGGLHCKP-- 284
DB 227 AMMKR--QFGSKV---PDVDALRKVNMVPMVTHPFDNRVPSVVOYLGIIHIDPAV 280
QY 285 ---AKPLPKEMEEFVQSSGNGVYVSLGSMV--SNTSEERANVIASALAKIPQVLMRF 339
DB 281 TSVADEIDNDIAELEFENS-TMGVYVYVSLGSSVASDMSNMLVVFETFRSIPRYLMKV 339
QY 340 DGNKP---DTLGLNTRLYKWIIPONDLLGHPKTKAFITGGNAGIYEAIYHGVPMVGVPIFG 397
DB 340 DKSDKIFDNIPIPSVILQRMFPQRRVLKRNKVKYFITGGVQSDEALIDAGVPMFGVPIFG 399
QY 398 DOLDNIAHMAKAGAAVEINKTISDILLALRTVITDSSYKENAMRLSIHHDDPKPL 457
DB 400 DQFNVYVMEYTYGIGGVVDLTIVDARQLTEIVMDVADNEKXYKNGITLWLRDAINDQPMRPL 459
```

OY 458 DRAVFWIRFVNRHKGK-KHLRSAAHDLTWFQHSID-VIGFLLTCVAT 503
DB 460 EKAVVYTHVARAKGKHLGTRAAVYTSKYAMFDLILPMLITFST 507

RESULT 15

US-09-370-838-36
; Sequence 36, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-36

Query Match 14.0%; Score 393; DB 4; Length 129;
Best Local Similarity 58.1%; Pred. No. 4.3e-33;

Matches 75; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

OY 303 GVVVFSLSGMSVNTSEERANVIASALAKIPQKVLNRFDPGNKPDITGLNTRLYKWIPONDL 362
DB 1 GVVVFSLSGMSVSEIPEKRAVAIADALGKIPQTVLWRYGTGRPSNLANNITLWQWLPONDL 60
OY 363 LGHPKTKAFITGCGNGGIYEATYHGVWVGVPIFGDOLDNIAHMKAGAAVEINFKMTS 422
DB 61 LGHPKTKAFITGASHGVNESICNGVPMVWIPFGDQMDNAKRRRTKGAGVTLNVLEMTS 120
OY 423 EDLLRLART 431
DB 121 EDLEDALKS 129

Search completed: December 5, 2003, 09:50:50
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 03:53:47 ; Search time 4104 Seconds

(without alignments)
16447.584 Million cell updates/sec

Title: US-09-980-729B-10

Perfect score: 1650

Sequence: 1 agcaactcggaacaagcat.....ccaattcaagaagacctg 1650

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

1: gb_ba:*

2: gb_ba:*

3: gb_ba:*

4: gb_ba:*

5: gb_ba:*

6: gb_ba:*

7: gb_ba:*

8: gb_ba:*

9: gb_ba:*

10: gb_ba:*

11: gb_ba:*

12: gb_ba:*

13: gb_ba:*

14: gb_ba:*

15: gb_ba:*

16: gb_ba:*

17: gb_ba:*

18: gb_ba:*

19: gb_ba:*

20: gb_ba:*

21: gb_ba:*

22: gb_ba:*

23: gb_ba:*

24: gb_ba:*

25: gb_ba:*

26: gb_ba:*

27: gb_ba:*

28: gb_ba:*

29: gb_ba:*

30: gb_ba:*

31: gb_ba:*

32: gb_ba:*

33: gb_ba:*

34: gb_ba:*

35: gb_ba:*

36: gb_ba:*

37: gb_ba:*

38: gb_ba:*

39: gb_ba:*

40: gb_ba:*

41: gb_ba:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1650	100.0	1650	6	AX040087
2	1444.4	87.5	2079	9	HAJ5162
3	1443.8	87.5	2129	9	BC026264
4	1433.2	86.9	2097	9	AF064200
5	1432.6	86.8	2093	6	AX411198
6	1432.6	86.8	2093	9	HSUDPCT
7	1431.6	86.8	2091	9	AF081793
8	1331.4	80.7	1639	6	AX548042
9	1292.4	78.3	2108	9	AF112112
10	1267.8	76.8	1766	9	BC030974
11	1254.6	76.0	1915	9	AF401657
12	1252.6	75.9	1855	6	AX336329
13	1252.6	75.9	1855	6	AX336696
14	1252.6	75.9	1855	6	AX409473
15	1252.6	75.9	1855	9	H0MDPCTA
16	1248	75.6	2648	9	MF091582
17	1238.6	75.1	1753	9	AF016310
18	1238.6	75.0	1722	9	AF016492
19	1236.8	75.0	1768	9	AF112113
20	1234.4	74.8	1662	9	AF294902
21	1232.8	74.7	1662	9	AF294901
22	1225.2	74.3	2799	6	AX411112
23	1225.2	74.3	2799	9	HSUGT2BIO
24	1208.6	73.2	1851	9	AF177272
25	1159.6	70.3	2107	6	AR168316
26	1159.6	70.3	2107	9	HSUS9209
27	1145.2	69.4	2114	9	AF180322
28	1133	68.7	2092	9	AF548389
29	1125.2	68.2	2090	6	AX410646
30	1125.2	68.2	2090	9	HSU08854
31	1115	67.6	2123	6	AX410640
32	1115	67.6	2123	9	HSU06641
33	1110.4	67.3	2075	9	AF072223
34	1002	60.7	1671	4	RABUGT2B
35	967.4	58.6	1636	6	AX327326
36	967.4	58.6	2759	6	AX395164
37	967.4	58.6	2974	6	AX376454
38	965.8	58.5	3006	6	AX497163
39	962.6	58.3	2818	4	OCU72742
40	962.2	58.3	1832	4	RABUGT2BA
41	956.8	58.0	1961	6	AX401727
42	956.8	58.0	1961	10	RATUDPCTP
43	955	57.9	2560	10	BC027200
44	947.4	57.4	1825	10	AB034987
45	936.8	56.8	1584	6	AX497165

ALIGNMENTS

RESULT 1	LOCUS	SEQUENCE	10	FROM	PATENT	MO0063351	DNA	linear	PAT	18-NOV-2000
AX040087	DEFINITION	Sequence 10 from Patent	MO0063351							
AX040087	ACCESSION	AX040087.1	GI:11230049							
AX040087.1	VERSION	AX040087.1	GI:11230049							
AX040087.1	KEYWORDS	AX040087.1	GI:11230049							
AX040087.1	SOURCE	AX040087.1	GI:11230049							
AX040087.1	ORGANISM	AX040087.1	GI:11230049							
AX040087.1	REFERENCE	AX040087.1	GI:11230049							
AX040087.1	AUTHORS	AX040087.1	GI:11230049							
AX040087.1	TITLE	AX040087.1	GI:11230049							
AX040087.1	JOURNAL	AX040087.1	GI:11230049							

Pred. No. is the number of results predicted by chance to have a

REFERENCE 2
AUTHORS Rittner, J. K., Chen, F., Sheen, Y. Y., Lubet, R. A., and Owens, I. S.
TITLE Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
log differences in activity toward parallel substrates including
hydroxycholeic acid and certain estrogen derivatives
JOURNAL Biochemistry 31 (13), 3409-3414 (1992)
MEDLINE 92207964
PUBMED 1554722
REFERENCE 3 (bases 1 to 2079)
AUTHORS Rittner, J. K.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-1998) Rittner J. K., Pharmacology and Toxicology,
Virginia Commonwealth University, P.O. Box 980613, Medical College
of Virginia, Richmond, Virginia, 23112-0613, USA
FEATURES
source location/qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UDPGT-1"
/clone_lib="Lambda bacteriophage gt11"
/tissue_lib="Liver"
1..2079
/gene="UGT2B4"
34..1620
/gene="UGT2B4"
/EC_number="2.4.1.17"
/function="transfers glucuronic acid to hydroxycholeic
acid and estrogen catechols"
/codon_start=1
/evidence="experimental"
/product="UDP-glucuronosyltransferase"
/protein_id="CAA06396.1"
/db_xref="GI:3135025"
/db_xref="SPTRMBL:O60867"
/translation="MSMKWTSALLLIQLSCYSSGSCGKVLWPTFEFSHMNNIKTIID
ELVORGEHVTVLASSASISFDNPSPTLKFVYVPSLKTFFEDIIKOLVRMAELPK
DTFWSYSONVOEITFNDILRKCKDIVSKKIMKLOESRPVVLADAVFPGELL
AELKIPVYSILRSPGAIIRKSGGLFPFSYIPVMSSELDQMTPIERKMMITVL
YFEFPOIFDMKKWDQYSEVLGRPTLSETMAADWLINWDFOPPHLLPNER
VGLHCKRPAKPLPEMEBFVQSGENGVSQSVNTSSEERANYIASLAKI POK
VLNRFDPDKPDLGLNTRLYKMI PONDILGHPKTRAFITGGANGIYEAIVHGIPMG
VPLPADOPNIAHMKAGAAVSLDFHTWSSTDLNLAKTIVINDPLYENAMKLSRIH
DQVPLDRAVFWIEFWARHKGAHLRAVADLWPFQHSIDLVTGFLACATVATFI
TKCLFCYWKFRTRGKGRD"
546
/gene="UGT2B4"
/citation=[1]
/replace="g"
548..549
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1187..1188
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1190
/gene="UGT2B4"
/citation=[1]
/replace="c"
1192
/gene="UGT2B4"
/citation=[1]
/replace="a"
1977..1978
/gene="UGT2B4"
/citation=[1]
/replace="cga"
2068
/gene="UGT2B4"
/citation=[1]
/replace="g"

BASE COUNT 634 a 396 c 436 g 613 t
ORIGIN
Query Match 87.5%; Score 1444.4; DB 9; Length 2079;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1521; Conservative 0; Mismatches 106; Indels 3; Gaps 1;
QY 2 GCAACTGGAAAAACAAGATTCATTTGATCATGATGTCATGAATGAGACTTCAGTCTT 61
DB 1 GCAACTGGAAAAACAAGATTCATTTGATCATGATGTCATGAATGAGACTTCAGTCTT 60
QY 62 CTCCTGATACAGCTGAGCTGTCTTACTTTAGCTCTGGAGTTGTGAAAGTGTGCTG 121
DB 61 CTGCTGATACAGCTGAGCTGTCTTACTTTAGCTCTGGAGTTGTGAAAGTGTGCTG 120
QY 122 CCCACAAATTCAGCCCTGATGTAATATAAGCAATCTCTGATGTAATCTTGCACAGA 181
DB 121 CCCACAAATTCAGCCCTGATGTAATATAAGCAATCTCTGATGTAATCTTGCACAGA 180
QY 182 GGTGATGAGGTAAGTGTATTTGGCATCTTCAGCTTCATTTCTTGCATCCCAAGCCCA 241
DB 181 GGTGATGAGGTAAGTGTATTTGGCATCTTCAGCTTCATTTCTTGCATCCCAAGCCCA 240
QY 242 TCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTTAACTGAAGTTGAGGATATT 301
DB 241 TCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTTAACTGAAGTTGAGGATATT 300
QY 302 ATCAAGAGCTGCTGTTAAGATGAGGCGAACTTCCAAAAGACATTTTGGTCAATTTT 361
DB 301 ATCAAGAGCTGCTGTTAAGATGAGGCGAACTTCCAAAAGACATTTTGGTCAATTTT 360
QY 362 TCACAAATACAGAAATCATGTGACATTTAATGACATCTTAGAAAGTTGTGTAAGAT 421
DB 361 TCACAAATACAGAAATCATGTGACATTTAATGACATCTTAGAAAGTTGTGTAAGAT 420
QY 422 ATAGTTCAATATAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCTT 481
DB 421 ATAGTTCAATATAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCTT 480
QY 482 GCAGATGCTGTTTCCCTTTGGTGAAGTGTGCGGCGAGTTACTTAAATACCTTTGTC 541
DB 481 GCAGATGCTGTTTCCCTTTGGTGAAGTGTGCGGCGAGTTACTTAAATACCTTTGTC 540
QY 542 TAGAGCTCGGCTCTCTCCGCTGCTACGAATTAAGAAAGCTAGTGGAGACTTCTGTC 601
DB 541 TAGAGCTCGGCTCTCTCCGCTGCTACGAATTAAGAAAGCTAGTGGAGACTTCTGTC 600
QY 602 CTTCTCTCTATGCTGCTGTTGTTATGTCAAGAACTAAGTCAAAATGACTTTGATAG 661
DB 601 CTTCTCTCTATGCTGCTGTTGTTATGTCAAGAACTAAGTCAAAATGACTTTGATAG 660
QY 662 AGGGTAAAAAATATGATCTATGCTTTATTTGAATTTTGGTCCAAATATTGACATG 721
DB 661 AGGGTAAAAAATATGATCTATGCTTTATTTGAATTTTGGTCCAAATATTGACATG 720
QY 722 AAGAAGTGGATCAGTCTACAGTGAAGTTCTAGAAAGCCCACTAGGTTATGAGACA 781
DB 721 AAGAAGTGGATCAGTCTACAGTGAAGTTCTAGAAAGCCCACTAGGTTATGAGACA 780
QY 782 ATGGCAAAAGTGCATATGCTTATTCGAAACTGAGTTTTCATTTTCTCACCACA 841
DB 781 ATGGCAAAAGTGCATATGCTTATTCGAAACTGAGTTTTCATTTTCTCACCACA 840
QY 842 CTCTTACCAATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCCAAACCCCTACCG 901
DB 841 CTCTTACCAATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCCAAACCCCTACCG 900
QY 902 AAGGAATGGAAGTTTGTCCAGAGCTGGAAGAAATGGTGTGTGTTTCTCTG 961
DB 901 AAGGAATGGAAGTTTGTCCAGAGCTGGAAGAAATGGTGTGTGTTTCTCTG 960
QY 962 GGGTCATGTCATTAACGTCAGAAAGAGGCAATGTAATGATCAGCCCTTGCC 1021
DB 962 GGGTCATGTCATTAACGTCAGAAAGAGGCAATGTAATGATCAGCCCTTGCC 1021

[illegible]

Db 371 TTCAAGTCAAGAATCATGTGGACATTATGACATACCTTAGAACTTCTGTAAAGA 430

421 TATAGTTCAATAGAGAACTTATGAGAACTACAGAGTCAAGATTGATGTTGTTCT 480
431 TATAGTTCAATAGAGAACTTATGAGAACTACAGAGTCAAGATTGATGTTGTTCT 490
481 TGCAGATGCTGTTTCCCTTTGGTGAAGTCTGCGAGTACTTAAATACCTTTGT 540
491 TGCAGATGCTGTTTCCCTTTGGTGAAGTCTGCGAGTACTTAAATACCTTTGT 550
541 CTACAGCTCCGCTTCTCTCTGCTGCTACGAAATGAAAGCATGTGAGGACTTCTGTT 600
551 CTACAGCTCCGCTTCTCTCTGCTGCTACGAAATGAAAGCATGTGAGGACTTCTGTT 610
601 CCTCTCTCTATGCTGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTTCAAGA 660
611 CCTCTCTCTATGCTGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTTCAAGA 670
661 GAGGTTAAAAATATGATCTATGCTTATTTGAAATTTGGTTCCAAATATTTGACAT 720
671 GAGGTTAAAAATATGATCTATGCTTATTTGAAATTTGGTTCCAAATATTTGACAT 730
721 GAAGAGTGGATCACTCTACAGTGAAGTTCTAGGAAGCCCACTAGCTATCTGAGAC 780
731 GAAGAGTGGATCACTCTACAGTGAAGTTCTAGGAAGCCCACTAGCTATCTGAGAC 790
781 AATGCAAAAGCTGACATATGCTTATTCGAACTACTGAGATTTTCAATTTCTCAACC 840
791 AATGCAAAAGCTGACATATGCTTATTCGAACTACTGAGATTTTCAATTTCTCAACC 850
841 ACTCTTACCAATGTTGAGTTGCTGAGAGCTCCACTGCAACCTGCAAAACCCCTAAC 900
851 ACTCTTACCAATGTTGAGTTGCTGAGAGCTCCACTGCAACCTGCAAAACCCCTAAC 910
901 GAAGAAATGGAAGTTGTCAGAGCTCGAGAAATGCTGTTGAGTTGTTCTCT 960
911 GAAGAAATGGAAGTTGTCAGAGCTCGAGAAATGCTGTTGAGTTGTTCTCT 970
961 GGGGTCATGCTGATGACATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
971 GGGGTCATGCTGATGACATGATGATGATGATGATGATGATGATGATGATGATGATG 1030
1021 CAAGATCCCAAAAGTTCTGTTGAGATTTGATGAGATTAACCATCTTTAGACT 1080
1031 CAAGATCCCAAAAGTTCTGTTGAGATTTGATGAGATTAACCATCTTTAGACT 1090
1081 CAATATCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
1091 CAATATCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
1141 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
1151 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1210
1201 TATGTTGGAGTTCCCATATTTGTTGATGATGATGATGATGATGATGATGATGATG 1260
1211 TATGTTGGAGTTCCCATATTTGTTGATGATGATGATGATGATGATGATGATGATG 1270
1261 AGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
1271 GGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1330
1321 GAGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
1331 GAGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1390
1381 CCATGATCAACCTGTAAGCCCTTATGATGATGATGATGATGATGATGATGATGATG 1440
1391 TCATGATCAACCTGTAAGCCCTTATGATGATGATGATGATGATGATGATGATGATG 1450
1441 CCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
1451 CCATGATCAACCTGTAAGCCCTTATGATGATGATGATGATGATGATGATGATGATG 1510
1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560

Db 1511 CTCTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1570
Qy 1561 AAAATGTTTATTTTCCGTCAAAATTTAAATTAAGATGAGAAAGAGGAGA 1620
Db 1571 AAAATGTTTATTTTCCGTCAAAATTTAAATTAAGATGAGAAAGAGGAGA 1627
Qy 1621 ATAGATCTTC 1631
Db 1628 TTAATTAAGTC 1638
RESULT 4
AF064200 2097 bp mRNA linear PRI 11-DEC-1998
LOCUS AF064200
DEFINITION Homo sapiens UDP-glucuronosyltransferase 2B4 precursor (UGT2B4)
ACCESSION AF064200
VERSION AF064200.1 GI:3153831
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Levesque, E., Beaulieu, M., Hum, M. and Belanger, A.
TITLE UDP-glucuronosyltransferase encoded by a polymorphic gene
JOURNAL PharmacoGenetics (1999) In press
AUTHORS Levesque, E., Beaulieu, M. and Belanger, A.
TITLE Direct Submission
JOURNALS Submitted (08-MAY-1998) Molecular Endocrinology, CHUL Research
Center, 2705 Laurier Blvd, Ste-Foy, PQ G1V 4G2, Canada
FEATURES
source
1..2097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2097
/gene="UGT2B4"
/allele="UGT2B4*E458"
26..1612
/gene="UGT2B4"
/note="E458 variant"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B4 precursor"
/protein_id="AAC95002.1"
/db_xref="GI:3153832"
/translation="MSMKWTSAALLLIQSCYFSSGCGKVLWMPTEFSHMNITILD
ELVORGEHVIVLSSAISFPDPSPTLKEEVVPSLTKEPFDILIKQVKNRAELPK
DIFMSYFSQVEIIMWTENDILKRCQIVSNKIMKMLQSRPDLVADAVFPGEILL
AEILKLPVYVSLRPSGVALERKSGGLPFPSPVVPVMSLSLQDMEIEVKNNIYVL
YFERFWQIFDMKRWDOFSEVLGRPTLSETMKADIMLRNWDFOFPHPLPNVFE
VGGHCHKPAKPLKREMEFVOSGENGVVFSISMSVNSSEERANYIASALAKIPOK
VLMRPDGNKPDITGLNRLRYKMI PONDILAHPTKPRATITGGAIGYEATYHGI PMVG
VPLRPDPNDIIMKAKGAAVSLDPHMSSTDILNAIKTYINPLVKNMKSRIH
DQPKPLRVAENVIEFVHRHKAHLVAHDLTFWPHSLDVTGFLACVATYIFII
TKCLPCWKRVRVIRGKGRD"

BASE COUNT 653 a 393 c 435 g 616 t
ORIGIN
Query Match 86.9%; Score 1433.2; DB 9; Length 2097;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 1511; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

Qy 10 AAAACAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69
Db 1 AAAACAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
Qy 70 ACAAGTGGCTGTTACTTGAAGTTGGAAGTGGTGGTGGCCACAGA 129
Db 61 ACAAGTGGCTGTTACTTGAAGTTGGAAGTGGTGGTGGCCACAGA 120

QY 181 AGGTCATGAGTGAAGTATGTCATCTGAGTTCATTTCTTTCGATCCCAAGAGCCC 240
 DB 184 AGGTCATGAGTGAAGTATGTCATCTGAGTTCATTTCTTTCGATCCCAAGAGCCC 243
 QY 241 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTAAAGCTGATTTGAGATAT 300
 DB 244 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTAAAGCTGATTTGAGATAT 303
 QY 301 TATTAACAGTGTGTTAAGATGAGGCAAACTTCCAAAGACATTTTGTATATTT 360
 DB 304 TATTAACAGTGTGTTAAGATGAGGCAAACTTCCAAAGACATTTTGTATATTT 363
 QY 361 TTCAAGATGACAAAGATCATGTGACATTTAATGACATTTGAAAGTCTGTAAGA 420
 DB 364 TTCAAGATGACAAAGATCATGTGACATTTAATGACATTTGAAAGTCTGTAAGA 423
 QY 421 TATAGTTCAAAATGAAAGCTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCT 480
 DB 424 TATAGTTCAAAATGAAAGCTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCT 483
 QY 481 TGCAGATGCTGTTTCCCTTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 484 TGCAGATGCTGTTTCCCTTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 QY 541 CTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 544 CTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
 QY 601 CCTCTCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 604 CCTCTCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
 QY 661 GAGGCTAAAGATGATCTATGCTGCTTATTTGAATTTGCTTCCAAATATTTGACAT 720
 DB 664 GAGGCTAAAGATGATCTATGCTGCTTATTTGAATTTGCTTCCAAATATTTGACAT 723
 QY 721 GAAGAAATGGGATGATCTATGCTGCTTATTTGAATTTGCTTCCAAATATTTGACAT 780
 DB 724 GAAGAAATGGGATGATCTATGCTGCTTATTTGAATTTGCTTCCAAATATTTGACAT 783
 QY 781 AATGGCAAAAGCTGACATATGCTTATTCGAACTAGTGGATTTTCAATTTCTCAACC 840
 DB 784 AATGGCAAAAGCTGACATATGCTTATTCGAACTAGTGGATTTTCAATTTCTCAACC 843
 QY 841 ACTCTTACCAATGTTGATGCTGTTGAGAGATCTCACTGCAAACTGCAAACTCCCTAACC 900
 DB 844 ACTCTTACCAATGTTGATGCTGTTGAGAGATCTCACTGCAAACTGCAAACTCCCTAACC 903
 QY 901 GAAGGAATGGAAGATGTTCCGACACTGCGAGAAATGCTGCTGCTGCTGCTGCTGCT 960
 DB 904 GAAGGAATGGAAGATGTTCCGACACTGCGAGAAATGCTGCTGCTGCTGCTGCTGCT 963
 QY 961 GGGGTGATGCTGATGACAGCTGAGAAAGAGGCAATGATTTGATGATGAGCTTGGC 1020
 DB 964 GGGGTGATGCTGATGACAGCTGAGAAAGAGGCAATGATTTGATGATGAGCTTGGC 1023
 QY 1021 CAAGATCCCAAAAGGTTCTGCTGAGATTTGATGGAATTAACCAATCTTTAGAGACT 1080
 DB 1024 CAAGATCCCAAAAGGTTCTGCTGAGATTTGATGGAATTAACCAATCTTTAGAGACT 1083
 QY 1081 CAATACCTCGCTGTAAGAGTGAATCCCAAGATGATCTTCTGCTGATCCCAAGAGAG 1140
 DB 1084 CAATACCTCGCTGTAAGAGTGAATCCCAAGATGATCTTCTGCTGATCCCAAGAGAG 1143
 QY 1141 AGCTTTTATCACTGATGATGGAATGAATGGATCTATGAAGCTATTTAACAATGGGCTCCC 1200
 DB 1144 AGCTTTTATCACTGATGATGGAATGAATGGATCTATGAAGCTATTTAACAATGGGCTCCC 1203
 QY 1201 TATGTTGGAGTTCCTCATTTTGTGATCAGCTTGATTAACATGATCTCAATGAAGGCCAA 1260
 DB 1204 TATGTTGGAGTTCCTCATTTTGTGATCAGCTTGATTAACATGATCTCAATGAAGGCCAA 1263

QY 1261 AGGAGCAGCTGTGTAAGAAATTAATCTTAACAACTATGACAGCAAGATTTACTGAGGCTTT 1320
 DB 1264 GGGAGAGCAGCTGTGTAAGAAATTAATCTTAACAACTATGACAGCAAGATTTACTGAGGCTTT 1323
 QY 1321 GAGAACAGCTATTAACGATCTCTTATTAAGAAATGCTATGATGATTAATCAAGATTTCA 1380
 DB 1324 GAGAACAGCTATTAACGATCTCTTATTAAGAAATGCTATGATGATTAATCAAGATTTCA 1383
 QY 1381 CCATGATCAACCTGTAAAGCCCTAGATGAGAGAGCTTGTGATGATGATGATGATGATGAT 1440
 DB 1384 TCATGATCAACCTGTAAAGCCCTAGATGAGAGAGCTTGTGATGATGATGATGATGATGAT 1443
 QY 1441 CCACAAAGAGCCAGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1444 CCATTAAGAGCCAGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1503
 QY 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 DB 1504 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563
 QY 1561 AAAATGTTTTTATTTTCTGCTCAAAATTAATTAAGAAAGTAAAGAAAGAGGGA 1620
 DB 1564 AAAATGTTTTTATTTTCTGCTCAAAATTAATTAAGAAAGTAAAGAAAGAGGGA 1620
 QY 1621 ATGATCTTTC 1631
 DB 1621 TTAATTAAGCTC 1631

RESULT 6
HSUPGT
DEFINITION Human mRNA for liver microsomal UDP-glucuronosyltransferase (UDPGT).
ACCESSION Y00317
VERSION Y00317.1 GI:37588
KEYWORDS transferase; UDP-glucuronosyltransferase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Jackson, M.R., McCarthy, L.R., Harding, D., Wilson, S., Coughtrie, M.W. and Burchell, B.
TITLE Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA
JOURNAL Biochem. J. 242 (2), 581-588 (1987)
MEDLINE 87241362
PUBMED 3109396
REFERENCE 2 (bases 1 to 2093)
AUTHORS Burchell, B.
TITLE Direct Submision
JOURNAL Submitted (26-JAN-1987) Burchell B., Dept. Biochemistry, Medical Sciences Institute, University Dundee, Scotland
FEATURES 1..2093
source location/Qualifiers
 1..2093
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HUG25"
 /tissue type="liver"
 /clone_lib="lambda gcl1 liver cDNA"
 38..1624
 /note="precursor"
 /codon_start=1
 /protein_id="CA68415.1"
 /db_xref="GI:37589"
 /translation="MSMKRTSALLIOLGCVPSGSGCKVLVMPTEESHMNTKITLD
 ELVORGHETVILASASISIFDPSPTLKEFVPSVLTKTEFDIITKOLVKRAELPK
 DTFWSYFSVOEIMWTENDILRFKCDIVSNKKLMDKESRDVLAADVFPFGEL
 AELIKLPVVRPSPGVAIEKSGGLPEPSPVPMVMSLSDQMTVFIEVKMIVL
 YFEFWFIQIDMKKMDQFYSFVLGRPTTLSETMAKADIWILIRNWDQFPHPLPNEF

OY	1	AGCAACTGGAAAAACAAGATTGCATTGGATCAGAGATGCTATGAATGAATGAGACTTCAGCTCT	60
Db	4	AGCAACTGGAAAAACAAGATTGCATTGGATCAGAGATGCTATGAATGAATGAGACTTCAGCTCT	63
OY	61	TCCTCGTAAACAGCTGAGCTGTACTTTAGCTCTGGAGTTGGAAAGTGTGCTGTG	120
Db	64	TCCTCGTAAACAGCTGAGCTGTACTTTAGCTCTGGAGTTGGAAAGTGTGCTGTG	123
OY	121	GCCCAACAATTCACGCCACTGGATGAATATATGAACAATCCCTGATGTAACCTGTCCAGAG	180
Db	124	GCCCAACAATTCACGCCACTGGATGAATATATGAACAATCCCTGATGTAACCTGTCCAGAG	183
OY	181	AGGTCATGAGGTGACTGTATTGGACATCTTCAGCTTCACATTTCTTTCGATATCCCAACAGCC	240
Db	184	AGGTCATGAGGTGACTGTATTGGACATCTTCAGCTTCACATTTCTTTCGATATCCCAACAGCC	243
OY	241	ATCTACTCTTAAATTTGAAATTATCTGTATCTTTAACTAAACCTGAGTTTGAGAGAT	300
Db	244	ATCTACTCTTAAATTTGAAATTATCTGTATCTTTAACTAAACCTGAGTTTGAGAGAT	303
OY	301	TATCAGAGAGCTGGTTAAGAGATGGGCGAAATTCGAAAGACATTTTGGTCATATTT	360
Db	304	TATCAGAGAGCTGGTTAAGAGATGGGCGAAATTCGAAAGACATTTTGGTCATATTT	363
OY	361	TTCAACAATACAAAGAAATCATGTGACATTTAATGACATCTTGAAGAAATTCGTAAAGA	420
Db	364	TTCAACAATACAAAGAAATCATGTGACATTTAATGACATCTTGAAGAAATTCGTAAAGA	423
OY	421	TATAGTTTCAATTAAGAAACTTATGAGAAACTACAGAGTCCAAAGTTTGATGTTGTTCT	480
Db	424	TATAGTTTCAATTAAGAAACTTATGAGAAACTACAGAGTCCAAAGTTTGATGTTGTTCT	483
OY	481	TGCAATGCTGTTTTCCCTTTGGTGAAGTGTGGCCGAGTACTTAAATATCCCTTGT	540
Db	484	TGCAATGCTGTTTTCCCTTTGGTGAAGTGTGGCCGAGTACTTAAATATCCCTTGT	543
OY	541	CTACAGCTCCGCTTCTCTCTCGGTAGCTACGCAATTTGAAAGCATAGTGAAGACTTCGT	600
Db	544	CTACAGCTCCGCTTCTCTCTCGGTAGCTACGCAATTTGAAAGCATAGTGAAGACTTCGT	603
OY	601	CCCTCTCTTCTATATGTCCTGTGTTATATGTCAGACCTAAGTACCAATGACTTTTATATGA	660
Db	604	CCCTCTCTTCTATATGTCCTGTGTTATATGTCAGACCTAAGTACCAATGACTTTTATATGA	663
OY	661	GAGGCTAAAAAAATATGATCTATGTGCTTATTTTGAATTTTGGTCCAAAATATTGACAT	720
Db	664	GAGGCTAAAAAAATATGATCTATGTGCTTATTTTGAATTTTGGTCCAAAATATTGACAT	723
OY	721	GAAGAAATGGGATTCAGTTCTACAGTGAAGTTCTAGAGAAACCCACTACGTTATCTGAGAC	780
Db	724	GAAGAAATGGGATTCAGTTCTACAGTGAAGTTCTAGAGAAACCCACTACGTTATCTGAGAC	783
OY	781	AATGGCAAAAGCTGACATATGGCTTATTTGAAAATCTAGGGATTTTCAATTTCTCTACCC	840

Db		784	AATGGCAAAAAGCTGACATTATGGCTTATTCCGAACACTCGGAAATTTCAAATTTCTCACC	843
OY		841	ACTTTTACCAAATGTGAGTTCTGTGGAGAGCTCCACTGCMAAACCTGCCAACCCTTAAC	900
Db		844	ACTCTTACCAAATGTTGATGATTCGTTTGAGAGACTCCACTGCMAAACCTGCCAACCCTTAAC	903
OY		901	GAAAGAAATGGAAGAGTTTGTCGAGACTCTCGAGGAAAAATGGTGTGGGTGTTTTCTCT	960
Db		904	GAAAGAAATGGAAGAGTTTGTCGAGACTCTCGAGGAAAAATGGTGTGGGTGTTTTCTCT	963
OY		961	GGGGTCGATGTCAGTAACAAGTCAGAAGAAAGGGCCAATGTAAATTCATCACGCCCTTGC	1020
Db		964	GGGGTCGATGTCAGTAACAAGTCAGAAGAAAGGGCCAATGTAAATTCATCACGCCCTTGC	1023
OY		1021	CAGATCCCACAAAAGGTTCTGTGAGATTTGATGGGAATPAAACCAATACTTTAGSACT	1080
Db		1024	CAGATCCCACAAAAGGTTCTGTGAGATTTGATGGGAATPAAACCAATACTTTAGSACT	1083
OY		1081	CAATACCTCGGCTGACAAAGGAGATACCCCAAGATGATCTTCTGGTATCCCAAAACCA	1140
Db		1084	CAATACCTCGGCTGACAAAGGAGATACCCCAAGATGATCTTCTGGTATCCCAAAACCA	1143
OY		1141	AGCTTTTATCATCTCATGSGTAGGAATGGAATGGGATCTATGAAGCTAATTAACATGGGACTCC	1200
Db		1144	AGCTTTTATCATCTCATGSGTAGGAATGGGATCTATGAAGCTAATTAACATGGGACTCC	1203
OY		1201	TATGCTGGAGATTCCCATPAITTTGCTGATCAGCTTGATPACATPAGCTCACATGAAGGCCA	1260
Db		1204	TATGCTGGAGATTCCCATPAITTTGCTGATCAGCTTGATPACATPAGCTCACATGAAGGCCA	1263
OY		1261	AGGAGCAGCTGTAAATAAATTAACCTCAAACTATGACAAAGGAAATTTTACGAGGGCTTT	1320
Db		1264	GGGAGCAGCTGTAGATTGCTGACCTTCAACACATGTCAATGACATPACTTACTCAATGCACT	1323
OY		1321	GAGAACAGTCATTACCGATTCCTTTAATAAGAAATGCTATGAGATTATCAAGAAATTA	1380
Db		1324	GAAACAGTAATTAATGATCCTTTAATAAGAAATGCTATGAGATTATCAAGAAATTA	1383
OY		1381	CCATGATCAACCTGTAAAGCCCCCTAGATCGACAGTCTTCTGATTCGAGTTGTCAATGCG	1440
Db		1384	TCATGATCAACCAATGAAAGCCCCCTAGATCGACAGTCTTCTGATTCGAGTTGTCAATGCG	1443
OY		1441	CCAAAGAGGCCAAGACCTGCGGATACAGCGGCCCATGACCTCACTGGTTCCAGACCTA	1500
Db		1444	CCATPAAAGSACCAGACCTTCCGGTTGACGCCCAAGACCTCACTGGTTCCAGATACA	1503
OY		1501	CTCTATAGATGTAGTTGGGTTCCCTGCTGACCTGTGTGGCAACTCTAATCTTGTGTCAC	1560
Db		1504	CTCTTGTGATGTACTGGGTTCCCTGCTGCTGTGTGTGGCAACTGTGATTTTCAATCATCAC	1563
OY		1561	AAAAATGTTTTTATTTCTCTGTCAAAAATTTAATPAAACTAGAAAGATAGAAAAGAGGA	1620
Db		1564	AAAAATGCTGTTTGTGTCTG---GAAGTTGTAGAACAGGAAAGAAAGGGAAGAGGA	1620
OY		1621	ATAGATCTTTC 1631	
Db		1621	TTAATTACGTC 1631	
<div style="text-align: center;">RESULT 7 AF081793</div>				
LOCUS				
DEFINITION Homo sapiens UDP glucuronosyltransferase 2B4 precursor (UGT2B4)				
ACCESSION AF081793				
VERSION AF081793.1 GI:3426331				
KEYWORDS				
SOURCE				
ORGANISM Homo sapiens (human)				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE				
1 (bases 1 to 2091)				

FEATURES	source
AUTHORS	Jin, C. J., Miners, J. O., Lillywhite, K. J. and Mackenzie, P. I.
TITLE	cDNA cloning and expression of two new members of the human liver UDP-glucuronosyltransferase 2B subfamily
JOURNAL	Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
MEDLINE	93326164
PUBMED	8333863
REFERENCE	2 (bases 1 to 2091)
AUTHORS	Mackenzie, P. I.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUL-1998) Clinical Pharmacology, Flinders University of South Australia, Bedford Park, SA 5042, Australia
FEATURES	Location/Qualifiers
source	1..2091
gene	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="4" 1..2091 /gene="UGT2B4" /note="formerly UGT2B11" /allele="UGT2B4*1i09, L396 variant" 36..1622 /gene="UGT2B4" /codon_start=1 /product="UDP glucuronosyltransferase 2B4 precursor" /protein_id="AAC32272.1" /db_xref="GI:3426332" /translation="MSMKMTSALLLIQLSCYFSSGSCGKVLWPFESHWNNIKTILD ELVORGEHVTLVASSASISPDNPSTLKFVYVPSLTKTFEDIILKOLVRYMLDPE DTPMSYLSVOEIMWTENDILRFCKDIOVSMLKTEQSRPVLVADAVFPGELL AELLKIPVYSLSRPGVAILKHSGLLPFSYLPVVMSELSDMTFIERKNNITYLV YFEFWFOIFDMKKNDQFYSEVLGRPFTLSETMAADVLINVWDPOFPHLLPVEVE VGGIACHKRAKPLPRKMEFVQSSGENGVVSLGSMVSNTESEERANYIASLAKIPOK VLWPFEDGKRPDTLLGKTRLYKMIIPONDILGHPKTRAFITGGANGIYEALVHGIPMGV VPLPADDOENIAHMKAKGAVALDPEHTMSSTDILNALKTIVINDPLVKNAMKLSRIHH DQPKRLDRAVFMVIEFWRHKGAGHLRAAMDILWPFQHSLSDVGFLLACATATYIFILL TKLCFCYWKFKRTGKKGRD"
BASE COUNT	638 a 401 c 437 g 615 t
ORIGIN	
Query Match	86.8%; Score 1431.6; DB 9; Length 2091;
Best Local Similarity	92.8%; Pred. No. 0;
Matches 1513; Conservative	0; Mismatches 114; Indels 3; Gaps 1;
QY	2 GCAACTGGAAAAACAAGCATTCGATTCGATCAGAGTGTATGAATAGTCACTTCAGCTTT 61
Db	3 GCAACTGGAAAAACAAGCATTCGATTCGATCAGAGTGTATGAATAGTCACTTCAGCTTT 62
QY	62 CTCTGTATACAGCTGAGCTGTTACTTTAGCTCTGCGAGTGTGGAAAGTCTGGTGTG 121
Db	63 CTGCTGTATACAGCTGAGCTGTTACTTTAGCTCTGCGAGTGTGGAAAGTCTGGTGTG 122
QY	122 CCCACAGAAATTCAGCACTGTAGTAAATTAAGAAACATCCGAGTGAACCTGTCCAGGA 181
Db	123 CCCACAGAAATTCAGCACTGTAGTAAATTAAGAAACATCCGAGTGAACCTGTCCAGGA 182
QY	182 GGTGATGAGTGATGATGATTTGGCATCTTCAGCTTCATTTCTTCGATCCCAACAGCCCA 241
Db	183 GGTGATGAGTGATGATGATTTGGCATCTTCAGCTTCATTTCTTCGATCCCAACAGCCCA 242
QY	242 TCTACTCTTAATTTGAAGTTTATCTGTATCTTTAACTAAAACTAGATTGAGATTT 301
Db	243 TCTACTCTTAATTTGAAGTTTATCTGTATCTTTAACTAAAACTAGATTGAGATTT 302
QY	302 ATCAAGCAGCTGTAAAGATAGGCAACAACCTCCAAAAGACACTTTTGGTCATATTTT 361
Db	303 ATCAAGCAGCTGTAAAGATAGGCAACAACCTCCAAAAGACACTTTTGGTCATATTTT 362
QY	362 TCACAAGTACAGAAGATATGATGTGACATTTATGATGATCTTAGAAATTTCTGTAAAGAT 421
Db	363 TCACAAGTACAGAAGATATGATGTGACATTTATGATGATCTTAGAAATTTCTGTAAAGAT 422
QY	422 ATAGTTTCAATTAAAGAACTTATGAGAAGAACTACAGAGTCAAGATTTGATGTGTCTT 481

Db	423	ATAGTTCAAAATTAAGAAACTTATATAGAACTACAGAGTCAGATTTGATGTTGTTCTT	482
Qy	482	GCAGATCGTGTTCCTCCCTTTGGAGCTGTCGCCGAGTACTTAAATAACCTTTGTC	541
Db	483	GCAGACGCTGTTTCCCTTTGGAGCTGTCGCCGAGTACTTAAATAACCTTTGTC	542
Qy	542	TACAGCCTCCGCTTCTCTCTGCGTACGCAATTGAAAAGCATAGTGGAGACTTCGTTC	601
Db	543	TACAGCCTCCGCTTCTCTCTGCGTACGCAATTGAAAAGCATAGTGGAGACTTCGTTC	602
Qy	602	CCTCCCTTCATGTCGCTGTGTGTAATGCAGAACTAAGTGCAGAAATACCTTCATAGAG	661
Db	603	CCTCCCTTCATGTCGCTGTGTGTAATGCAGAACTAAGTGCAGAAATACCTTCATAGAG	662
Qy	662	AGGGTAAAAAATATGATCTATGTGCTTTATTTTGAATTTTGTTCCAAATAATTGACATG	721
Db	663	AGGGTAAAAAATATGATCTATGTGCTTTATTTTGAATTTTGTTCCAAATAATTGACATG	722
Qy	722	AAGAAATGGGATCACTTCTACAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGACA	781
Db	723	AAGAAATGGGATCACTTCTACAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGACA	782
Qy	782	ATGGCAAAAGGTGATATGAGCTAATTCGAAACTACAGTGGGATTTTCAATTCTCCACCCA	841
Db	783	ATGGCAAAAGGTGATATGAGCTAATTCGAAACTACAGTGGGATTTTCAATTCTCCACCCA	842
Qy	842	CTCTTACCAATGTGTGAGTTCGTTGGAGAGACTCCACTGCAGAACTGCGAAACCCCTACCG	901
Db	843	CTCTTACCAATGTGTGAGTTCGTTGGAGAGACTCCACTGCAGAACTGCGAAACCCCTACCG	902
Qy	902	AAGGAAATGGAAGAATTGTTCAGAGCTCTGGAGAAATAGTGTGTGTGTTCTCTG	961
Db	903	AAGGAAATGGAAGAATTGTTCAGAGCTCTGGAGAAATAGTGTGTGTGTTCTCTG	962
Qy	962	GGGTGCATGTGCAGCTACGTCAGAGAAAGGGCCAAATGTATTTGCATCAGCCCTTGCC	1021
Db	963	GGGTGCATGTGCAGCTACGTCAGAGAAAGGGCCAAATGTATTTGCATCAGCCCTTGCC	1022
Qy	1022	AAGATCCCAAAAAGTTCTGTGGAGATTTGATGGGAAATAAACAGATACTTTAGACATC	1081
Db	1023	AAGATCCCAAAAAGTTCTGTGGAGATTTGATGGGAAATAAACAGATACTTTAGACATC	1082
Qy	1082	AATATCTGGGCTGTACAAATGAGATACCCAGAAATATCTTGTGTATCCCAAAACCAA	1141
Db	1083	AATATCTGGGCTGTGTACAAATGAGATACCCAGAAATATCTTGTGTATCCCAAAACCAA	1142
Qy	1142	GCTTTTATCACTCATGTGTGAATGAATGGGATCTATGAAGTATTTTACATGGGGTCCCT	1201
Db	1143	GCTTTTATCACTCATGTGTGTGAAGCCAAATGGCAATCTATGAGGCAATCTACATGGAATCCCT	1202
Qy	1202	ATGTGTGGAGATTCCTCATATTTTGTGTGATCAGGTTGATTAACATAGCTGCACATAGGCCCAA	1261
Db	1203	ATGTGTGGAGATTCCTCATATTTTGTGTGATCAGGTTGATTAACATAGCTGCACATAGGCCCAAG	1262
Qy	1262	GGAGCAGCTGTGAGAAATTAACCTTCAAACTATGACAAAGCGAAGATTTTACTGAGGGCTTTG	1321
Db	1263	GGAGCAGCTGTGTAGTTGAGACTTCACACCAATGTGAGTACAGACTTACTCAATGACATG	1322
Qy	1322	AGAAACGTCACTTACCGATTCTCTTATTAAGAAGATGTATGAGATTATCAAGATTTAC	1381
Db	1323	AAGACAGTATTAATGATCTCTTTATTAAGAAGATGTATGAGAAATTAATCAAGAAATTCAT	1382
Qy	1382	CATATATCAACCTGTAAAGCCCTGTAGTCGAGCACTCTTGTGATTCGAGTTTGTATGTCGC	1441
Db	1383	CATATATCAACAGTGAAGCCCTTGTAGTCGAGCACTCTTGTGATTTGAATTTGTCAATGCGC	1442
Qy	1442	CACAAAGAGCCAAAGCACTTGCGATACAGCTGCCATGACCTCACTGCTTCCAGACATAC	1501
Db	1443	CATTAAGAGAGCAAGCACTTCCGGGTGACAGCCACAGCCTCACTGCTTCCAGATACAC	1502
Qy	1502	TCTATAGATGATGTGGGTTCTGTGACCTGTGTGCACTGCTATATTTCTTGTTCACA	1561

Db 1503 TCCTTGATGTGACTGGGTCTCTGCTGCGCTGTGTGCACTGTGATTCATCATCA 1562
Qy 1562 AAATGTTTTTATTTTCTCTGTCAAAATTAATAAATAGATGAAAGGGAA 1621
Db 1563 AAATGTCGTTTTGTGCTG--GAAGTTGTAGAACAGAGAAAGAGGAGAT 1619
Qy 1622 TAGATCTTTC 1631
Db 1620 TAATTACCTC 1629

RESULT 8
AX548042 1639 bp DNA Linear PAT 26-NOV-2002
LOCUS Sequence 18 from Patent WO0206654.
DEFINITION AX548042
ACCESSION AX548042.1 GI:25813138
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1
Astromoff, A., Au-Young, J., Baughn, M.R., Ding, L., Duggan, B.M.,
Foreyshe, I.J., Gietzen, K.J., Griffin, J.A., Lee, E.A., Lu, Y.,
Richardson, T.W., Ring, H.Z., Sanjanwala, M.M., Swarnakar, A.,
Walia, N.K., Warren, B.A., Xu, Y., Yue, H. and Zebberjadian, Y.
TITLE Drug metabolizing enzymes
JOURNAL Patent: WO 0206654-A 18 29-AUG-2002;
Incyte Genomics, Inc. (US)
FEATURES
SOURCE 1. 1639
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7493833CB1"
BASE COUNT 470 a 316 c 372 g 481 t
ORIGIN

Query Match 80.7%; Score 1331.4; DB 6; Length 1639;
Best Local Similarity 89.6%; Pred. No. 0;
Matches 1431; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

Qy 35 ATGTATGAATGAGCTTCACTCTCTCTCTGATATACAGCTGAGCTGTTACTTACTCT 94
Db 1 ATGTATGAATGAGCTTCACTCTCTCTCTGATATACAGCTGAGCTGTTACTTACTCT 60
Qy 95 GGGAGTTGTGAAAAGTGTGTGTGGCCCAAGAAATTCAGCACTGATGAATTAAG 154
Db 61 GGGAGTTGTGAAAAGTGTGTGTGGCCCAAGAAATTCAGCACTGATGAATTAAG 120
Qy 155 ACAATCTCGATGAACCTGTCCAGAGAGTCAAGAGTCACTGATGAATTAAG 214
Db 121 ACAATCTCGATGAACCTGTCCAGAGAGTCAAGAGTCACTGATGAATTAAG 180
Qy 215 TCCATTTCTTGTGATCCCAAGAGCCATCTACTCTTAATTTGAAGTTATCTGATCT 274
Db 181 TCCATTTCTTGTGATCCCAAGAGCCATCTACTCTTAATTTGAAGTTATCTGATCT 240
Qy 275 TTAATAAACTGAGTTGAGATATTAACAAGACAGCTGTTAAGAGTGGCAGAACTT 334
Db 241 TTAATAAACTGAGTTGAGATATTAACAAGACAGCTGTTAAGAGTGGCAGAACTT 300
Qy 335 CCAAAAAGACACTTTTGTGATATTTTCAACAAGTCAAGAAATCATGTGCAATTAT 394
Db 301 CCAAAAAGACACTTTTGTGATATTTTCAACAAGTCAAGAAATCATGTGCAATTAT 360
Qy 395 GACATCTAGAAAGTCTGTAAGATATAGTTCAATTAAGAACTTAAGAAAGTCA 454
Db 361 GACATCTAGAAAGTCTGTAAGATATAGTTCAATTAAGAACTTAAGAAAGTCA 420
Qy 455 CAGAGTCAAGATTTGATGTTGTTCTTGAGAGTGTCTTCCCTTGGAGGTGCTG 514
Db 421 CAGAGTCAAGATTTGATGTTGTTGTTGAGAGTGTCTTCCCTTGGAGGTGCTG 480

Qy 515 GCGAGTTACTTAATATACCTTTGTGTCAAGCTCGGCTCTCTCTGAGTCAAGAT 574
Db 481 GCTAGCTATTTTAACATACCTTTGTGTCAAGCTCGGCTCTCTCTGAGTCAAGAT 540
Qy 575 GAAAGCATAGTGAAGACTTCTGTCCCTCTCTCTGATGCTGTTGTTATGAGAA 634
Db 541 GAAAGCATAGTGAAGACTTCTGTCCCTCTCTCTGATGCTGTTGTTATGAGAA 600
Qy 635 CTAGTGAACCAATGACTTTTATAGAGGTTAAATAATGATCTATGCTTATTTT 694
Db 601 TTAATGATCAAAAGACTTTTATGAGAGGTTAAATAATGATCTATGCTTATTTT 660
Qy 695 GAATTTGGTCCAAATATTTGACATGAGAGTGGGATGCTCTACAGTGAAGTTCTA 754
Db 661 GACTTTGGTCCAAATATTTGACATGAGAGTGGGATGCTCTATAGGAACTCTA 720
Qy 755 GGAAGACCACTAGCTTATCTGAGCAATGGAAGCTGACATAGGCTTATTCGAAAC 814
Db 721 GGAAGACCACTAGCTTATCTGAGCAATGGAAGCTGACATAGGCTTATTCGAAAC 780
Qy 815 TACTGGATTTTCAATTTCTCTCAACCCTCTTACCAATGTTGAGTCTGTGAGACTC 874
Db 781 TCTGGAATTTTCAATTTCTCTCAACCCTCTTACCAATGTTGAGTCTGTGAGACTC 840
Qy 875 CACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGTGTTCAGAGCTTGA 934
Db 841 CACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGTGTTCAGAGCTTGA 900
Qy 935 GAAATGTTGTGTGTGTGTGTCTCTGTGGGATGATGTCAGTAAACAGCTGAGAAAG 994
Db 901 GAAATGTTGTGTGTGTGTGTCTCTGTGGGATGATGTCAGTAAACAGCTGAGAAAG 960
Qy 995 GCCAATGATTTGATCAAGCTTGTCCCAAGTCCCAAGAAAGTCTCTGAGAAAGTTGAT 1054
Db 961 GCCAATGATTTGATCAAGCTTGTCCCAAGTCCCAAGAAAGTCTCTGAGAAAGTTGAT 1020
Qy 1055 GGAATTAACAGATATCTTGAAGTCAATCTGAGTCAAGTGAATCCCAAGAT 1114
Db 1021 GGAATTAACAGATATCTTGAAGTCAATCTGAGTCAAGTGAATCCCAAGAT 1080
Qy 1115 GATCTTCTGTGTCAATCCCAAGCAAGCTTTTATCACTCATGTGTGAATGAGATC 1174
Db 1081 GATCTTCTGTGTCAATCCCAAGCAAGCTTTTATCACTCATGTGTGAATGAGATC 1140
Qy 1175 TATGAAGTATTTACATGGGATCCCTATGTTGGGAGTCCCATATTTGGATCACTT 1234
Db 1141 TATGAAGTATTTACATGGGATCCCTATGTTGGGAGTCCCATATTTGGATCACTT 1200
Qy 1235 GATTAACATAGCTCAATGAAGGCAAGAGGAGCTGTAAGTAACTTCAAACTATG 1294
Db 1201 GATTAACATAGCTCAATGAAGGCAAGAGGAGCTGTAAGTAACTTCAAACTATG 1260
Qy 1295 ACAAGCGAAGATTTACTGAGGGCTTGAAGAACATCTTACGATTCCTCTTATTAAGAG 1354
Db 1261 TCGAGTACAGCTTGTGTAAGTGAAGAGATTAATTAATCTTCAATTAAGAG 1320
Qy 1355 AATGCTATGAATTTCAAGAAATTCACATGATCAACCTGTAAGGCCCCCTGATGAGCA 1414
Db 1321 AATGCTATGAATTTCAAGAAATTCACATGATCAACCTGTAAGGCCCCCTGATGAGCA 1380
Qy 1415 GTCTTCTGATGAGTTGTGATGAGGCAAGAGGCAAGAGCTGATGAGCTGCC 1474
Db 1381 GTCTTCTGATGAGTTGTGATGAGGCAAGAGGCAAGAGCTGATGAGCTGCC 1440
Qy 1475 CATGACCTCACTGTTTCAAGCACTCTATGATGATGAGTGGTCTCTGCTGACTGT 1534
Db 1441 CACGACCTCACTGTTTCAAGCACTCTTATGATGATGAGTGGTCTCTGCTGACTGT 1500
Qy 1535 GTGCACTGCTATATTTCTTGTCAAAAGTTTTATTTCTCTGTCAAAATTTAT 1594
Db 1501 GTGCACTGCTATATTTCTTGTCAAAAGTTTTATTTCTCTGTCAAAATTTAT 1560

Qy 1595 AAAAGTAAAGATAGAAAGAGGAAATAGATCTTTC 1631
Db 1561 AGAACAGAAAGAGGAGAAAGATTAATTACGTC 1597

RESULT 9
AF112112 2108 bp mRNA linear PRI 11-APR-1999
LOCUS Macaca fascicularis UDP-glucuronosyltransferase 2B19 precursor,
DEFINITION mRNA, complete cds.
ACCESSION AF112112
VERSION AF112112.1 GI:4580601
KEYWORDS Macaca fascicularis (craab-eating macaque)
SOURCE Macaca fascicularis
ORGANISM Bukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
REFERENCE 1 (bases 1 to 2108)
AUTHORS Belanger, G., Barbier, O., Hum, D.W. and Belanger, A.
TITLE Molecular cloning, expression and characterization of a monkey
steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates
testosterone
JOURNAL Eur. J. Biochem. 260 (3), 701-708 (1999)
MEDLINE 99203465
PUBMED 10102998
REFERENCE 2 (bases 1 to 2108)
AUTHORS Belanger, G., Barbier, O., Hum, D.W. and Belanger, A.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada

FEATURES
source
1..2108 location/Qualifiers
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
26..1612
/EC_number="2.4.1.17"
/note="UGT2B19"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B19 precursor"
/protein_id="AAD2435.1"
/db_xref="GI:4580602"
/translation="MSMKWTSALLLIOLSCYLSFGSGKULWPTESHMNTITLD
ELVORGHEVTLAYSTILDPNNPSPLKEICTPSLLETFQDSVQLVKRMSDIKK
DTFMPHFLHVEEMWTYGDMLRKECKOVNSKMLKMOESRFVVLADALSPEGELI
AELIKIPFVVSILRSPGVALEKHGGGELFPSPYVPMSELRDPTMERVONMIYVY
YDFEWMVDVKNMDOFYSKVLGRPTLFEIMAKAELIMLENDYDFQEPHLLPNEVE
VGGIHCAPKAPKPEMEEPVSGDNGVVSLSGMSVNSSEEPANVIAASLAKIPK
VLRFDPENKPDITGLANTOLYKMLPONDILGHPKTRAPITHEGANGIYBAIHGIPMG
VPLFADQPDNIAMHAKGAARLDPDMSSTDILNAKTVIDIYENAKKSSIH
DQVPKPLDRAVFWIEFWRHKGAGHLRVAADHLTFWYHSLDIVIGFLACVATVIFII
TKCLFCVWKEFVTRKKGRD"

BASE COUNT 649 a 393 c 438 g 628 t
ORIGIN

Query Match 78.3%; Score 1292.4; DB 9; Length 2108;
Best Local Similarity 87.7%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

Qy 190 GGTGACTGATATGGCATCTTCAGCTTCATTTCTTGATCCACAGCCCATCTACT 249
Db 181 GGTGACTGATCTGGCATATTCATTCATTTCTTGATCCACAGCCCATCTACT 240
Qy 250 TAAATTTGAAGTTTATCTGATCTTTAACTAACTGAGTTGAGATATTAACAGCA 309
Db 241 TAAATTTGAAGTTTATCTGATCTTTAACTAACTGAGTTGAGATATTAACAGCA 300
Qy 310 GGTGGTTAAGAGATGGGACAGACTTCAGAAAGACATTTTGGTCATATTTTACAGT 369
Db 301 ACTGGTTAAGAGATGGGACAGACTTCAGAAAGACATTTTGGGACATTTTACAGT 360
Qy 370 ACAAGAAATCATGTGACATTTAATGACATATAGAAAGTTCTGTAAGATATAGTTTC 429
Db 361 ACAAGAAATCATGTGACATATGAGTGCATGATTAAGAAAGTTCTGTAAGATATGTTTC 420
Qy 430 AATATGAATCTATGAGAAACTACAGAGTCAAGATTTGATGTTCTTGACAGATGC 489
Db 421 TAAATGAATCTATGAGAAACTACAGAGTCAAGATTTGATGTTCTTGCGGATGC 480
Qy 490 TGTTCCTCCCTTGGTGAAGTGTGGCCGAGTTAATAAATCCCTTGTCTACAGCT 549
Db 481 TATTCCTCCCTTGGTGAAGTGTGGCCGAGTTAATAAATCCCTTGTCTACAGCT 540
Qy 550 CCGCTTCTCTCGGCTACGCAATTTGAAAGACATAGTGAAGACTTCTGTTCCCTCTTC 609
Db 541 CCGCTTCTCTCGGCTACGCAATTTGAAAGACATAGTGAAGACTTCTGTTCCCTCTTC 600
Qy 610 CTATGTCCTGTTGTTATGTCAGAACTAAGTCAACAAATGACTTTCATAGAGAGGTAA 669
Db 601 CTATGTCCTGTTGTTATGTCAGAACTAAGTCAACAAATGACTTTCATAGAGAGGTAA 660
Qy 670 AATATATCTATGCTGTTTATTTTGAATTTTGGTCCAAATATTTACATGAAGATG 729
Db 661 AATATATCTATGCTGTTTATTTTGAATTTTGGTCCAAATATTTACATGAAGATG 720
Qy 730 GATCATGTTCTACAGTAAGTTCTAGAGAACCCACTACTGATCTGAGCAATAGGCAAA 789
Db 721 GATCATGTTCTACAGTAAGTTCTAGAGAACCCACTACTGATCTGAGCAATAGGCAAA 780
Qy 790 AGCTGACATATGCTTATTTGAAACTACAGGATTTTCAATTTCTGACCCACTCTTACC 849
Db 781 AGCTGACATATGCTTATTTGAAACTACAGGATTTTCAATTTCTGACCCACTCTTACC 840
Qy 850 AATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCGCAAACTCCAGAGAAAT 909
Db 841 AATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCGCAAACTCCAGAGAAAT 900
Qy 910 GGAAGAGTTGTCAGAGCTCTGAGAAATGATGTTGTTCTCTGGGGTGCAT 969
Db 901 GGAAGAGTTGTCAGAGCTCTGAGAAATGATGTTGTTCTCTGGGGTGCAT 960
Qy 970 GGTGATGACAGCTCAGAGAAAGGCGCAATGTAATTCATGACCCCTTGCAAGATCCC 1029
Db 961 GGTGATGACAGCTCAGAGAAAGGCGCAATGTAATTCATGACCCCTTGCAAGATCCC 1020
Qy 1030 ACAAAAGTTCTGTGAGATTTGATGAGATTAACCAAGATCTTAGAGATCTCAATCTCG 1089
Db 1021 ACAAAAGTTCTGTGAGATTTGATGAGATTAACCAAGATCTTAGAGATCTCAATCTCA 1080
Qy 1090 GGTGATGACAGTGAATCCAGATGATCTTCTGTCATCCCAAAACCAAGGTTTAT 1149
Db 1081 GGTGATGACAGTGAATCCAGATGATCTTCTGTCATCCCAAAACCAAGGTTTAT 1140
Qy 1150 CACTCATGTGATGATGATGAGATCTATGAAGCTATTTACATGAGGTTCCCTATGATGAG 1209
Db 1141 AACTCATGATGATGATGAGATCTATGAAGCTATTTACATGAGATGATGATGAG 1200
Qy 1210 AGTTCCCATATTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1269
Db 1201 CGTTCCATATTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
Qy 1270 TGTAGAAATTAATCTTCAAACTATGACAGCAAGATTTTCTGAGGGCTTGAAGACAGT 1329

Db 1261 TGTAGACTGAGCTTCGACACAAATGTGAGTACTTCAATGCACTGAGAACAGCT 1320
Qy 1330 CATTAACGATTCCTCTTTAAAGAGATGCTATGAGATTATCAAGATTCACTGATCA 1389
Db 1321 AATAAATGATCTCTATATTAAGAAAGTATGATAATATCAAGATTATCAATGATCA 1380
Qy 1390 ACCTGTAAAGCCCTAGATGAGAGCTCTTCTGATGAGTTTGTGATGCGCCCAAAAG 1449
Db 1381 ACCAGTGAAGCCCTGAGATGAGAGCTCTTCTGATGAGTTTGTGATGCGCCCAAAAG 1440
Qy 1450 AGCCAAAGACCTGCGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGT 1509
Db 1441 AGCCAAAGACCTGCGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGT 1500
Qy 1510 TGTGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGT 1569
Db 1501 TGTGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGTGCCTGATGAGT 1560
Qy 1570 TTTATTTTCTGTAATAATTTAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1629
Db 1561 GTTTTGTCTG--GAACTTTGTAAGAACAAAGAAAGAAAGAAAGTAAATCA 1617
Qy 1630 TC 1631
Db 1618 TC 1619

RESULT 10

BC030974 1766 bp mRNA linear PRI 13-JUN-2002
LOCUS Homo sapiens, UDP glycosyltransferase 2 family, polypeptide B7,
DEFINITION clone MGC:32620 IMAGE:4593426, mRNA, complete cds.

ACCESSION BC030974
VERSION BC030974.1 GI:21411301
KEYWORDS MGC.

SOURCE Homo sapiens (human)
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (base 1 to 1766)

REFERENCE 1 Strausberg, R.
AUTHORS Direct Submission
TITLE Submitted (03-JUN-2002) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrived by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdedpax11.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

FEATURES
source
1..1766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/db_xref="taxon:9606"
/clone="MGC:32620 IMAGE:4593426"
/tissue_type="kidney"

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
Series: IRAL Plate: 41 Row: e Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4507824.
Location/Qualifiers

CDS
/clone_1lb="NIH_MGC_75"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
47..1636
/codon_start=1
/product="UDP glycosyltransferase 2 family, polypeptide
B7"

BASE COUNT 546 a 327 c 381 g 512 t
ORIGIN
Query Match 76.8%; Score 1267.8; DB 9; Length 1766;
Best Local Similarity 86.1%; Pred. No. 2e-313;
Matches 1404; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 1 AGCAACTGGAAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 13 AGCAACTGGAAAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 72
Qy 61 TCTCTGATACAGCTGAGCTGTTTCTTAACTGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 120
Db 73 TTTCTAATACAACTGAGCTTTTCTTAACTGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 132
Qy 121 GCCCAAGATTCAGCCATGAGTGAATATTAACAATCTGATGATGATGATGATGATGATGAT 180
Db 133 GGCAGCAAGATTCAGCCATGAGTGAATATTAACAATCTGATGATGATGATGATGATGATGAT 192
Qy 181 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 193 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
Qy 241 ATCTACTCTTAAATTTGAAGTTTATCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 253 ATCTACTCTTAAATTTGAAGTTTATCTGATGATGATGATGATGATGATGATGATGATGAT 312
Qy 301 TATCAAGAGCTGTTAAGATGAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 313 CATCATGCAACAGATTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
Qy 361 TTCAAGATGCAAGAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 373 TTCAAGATGCAAGAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
Qy 421 TATAGTTTCAATTAAGAAATTTAAGAAATTTAAGAAATTTAAGAAATTTAAGAAATTTAAG 480
Db 433 TGTAGTTTCAATTAAGAAATTTAAGAAATTTAAGAAATTTAAGAAATTTAAGAAATTTAAG 492
Qy 481 TGCAGATGCTTTTCCCTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 540
Db 493 TGCAGATGCTTTTCCCTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 552
Qy 541 CTACAGCTGCGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 553 GTACAGCTGCGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
Qy 601 CCTCTCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 613 CCTCTCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
Qy 661 GAGGTAATAAATATGATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 673 GAGGTAATAAATATGATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 732

QY 721 GAAGAAGTGGATGCTTCTACAGTGAAGTCTAGGAAGACCCTAGCTATCTGAGAC 780
 DB 733 GAAGAAGTGGATGCTTCTACAGTGAAGTCTAGGAAGACCCTAGCTATCTGAGAC 792
 QY 781 AATGGCAAAAGCTCAATATGCTTATTCGAACCTACTGGGATTTTCAATTTCTCAACC 840
 DB 793 AATGGGAAAAGCTGACGATATGCTTATTCGAACCTCTGGAAATTTTCAATTTCCATATTC 852
 QY 841 ACTCTTACCAATGTTGAGTTCGTGAGAGCTCCACTGCAACCTGCAAAACCCCTGACC 900
 DB 853 ACTCTTACCAATGTTGAGTTCGTGAGAGCTCCACTGCAACCTGCAAAACCCCTGACC 912
 QY 901 GAAGAAGTGGATGCTTCTACAGTGAAGTCTAGGAAGACCCTAGCTATCTGAGAC 960
 DB 913 TAAGGAATGGAACACTTTGACAGAGCTCTGAGAAAATGCTGTGCTGCTTCTCT 972
 QY 961 GGGGTGATGCTGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1020
 DB 973 GGGGTGATGCTGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1032
 QY 1021 CAAGATCCCAAAAGTTCGTGAGAGTTCGTGAGAGTTCGTGAGAGTTCGTGAGAGTTCGTGAG 1080
 DB 1033 CAAGATCCCAAAAGTTCGTGAGAGTTCGTGAGAGTTCGTGAGAGTTCGTGAGAGTTCGTGAG 1092
 QY 1081 CAATACTCGGCTGATCAAGTGGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1140
 DB 1093 CAATACTCGGCTGATCAAGTGGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1152
 QY 1141 AGCTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 DB 1153 AGCTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212
 QY 1201 TATGCTGGGAGTTCCTATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1213 TATGCTGGGAGTTCCTATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1272
 QY 1261 AGGAGAGCTGTAGAAATTAACCTCAAAAATGATGACGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1273 GGGAGAGCTGTAGAAATTAACCTCAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1332
 QY 1321 GAGAAGAGCTGATCCGATTCCTCTTATTAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 DB 1333 GAGAAGAGCTGATCCGATTCCTCTTATTAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1392
 QY 1381 CCATGATCAACCTGATTAAGCCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1393 ACATGATCAACCTGATTAAGCCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1452
 QY 1441 CCAGAAAGAGCCAGAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1453 CCAGAAAGAGCCAGAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1512
 QY 1501 CTCTATGAT 1560
 DB 1513 CTCTATGAT 1572
 QY 1561 AAAATGTTTTTATTTCTCTGCTCAAAAATTTAATTAAGATGATGATGATGATGATGATGATGATGAT 1620
 DB 1573 AAAATGTTTTTATTTCTCTGCTCAAAAATTTAATTAAGATGATGATGATGATGATGATGATGATGAT 1632
 QY 1621 ATAGATCTTTC 1631
 DB 1633 TTAGTTATATC 1643

RESULT 11
 AF401657 1915 bp mRNA linear PRI 11-JUL-2002
 LOCUS AF401657
 DEFINITION Macaca fascicularis UDP-glucuronosyltransferase 2B30 (UGT2B30)
 ACCESSION AF401657
 VERSION AF401657.1 GI:18139799

KEYWORDS
 SOURCE Macaca fascicularis (crab-eating macaque)
 ORGANISM Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.
 REFERENCE
 AUTHORS 1 (bases 1 to 1915)
 TITLE Girard, C., Bardier, O., Turgeon, D. and Belanger, A.
 Isolation and characterization of the monkey UGT2B30 gene that
 encodes a uridine diphosphate-glucuronosyltransferase enzyme active
 on mineralocorticoid, glucocorticoid, androgen and oestrogen
 hormones
 JOURNAL Biochem. J. 365 (Pt 1), 213-222 (2002)
 MEDLINE 22067079
 PUBMED 12071853
 REFERENCE 2 (bases 1 to 1915)
 AUTHORS Girard, C., Bardier, O. and Belanger, A.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-2001) Molecular Endocrinology, CHUL Research
 Center, 2705 Laurier Blvd., Sainte-Foy, QC G1V 4G2, Canada
 FEATURES
 source
 1..1915
 /organism="Macaca fascicularis"
 /mol_type="mRNA"
 /db_xref="taxon:9541"
 1..1915
 /gene="UGT2B30"
 1..1587
 /gene="UGT2B30"
 /codon_start=1
 /product="UDP-glucuronosyltransferase 2B30"
 /protein_id="AAL60145.1"
 /db_xref="GI:18139800"
 /translation="MSMKRTSALLLLQLSCLSSGCKGLVWPTFSHMMNKTILD
 ELVORGEVAVLVAVSPILPGPNPSPALKEIPSLTBTEFSDSYQLVKRSDLPK
 GTRDPHPLQVQEMMTYMDIRKFCQDVSNKMLKLDQSRDVLADLISGEBLL
 AELIKLPFVYSLEFSPGYAIEKRGDFLPSPVPMSEFSDQMTFMEVKMIVY
 YDFPFOAMDTKMDQFSEVLGRPTTLFETNAKAEIWLIRNWDQFPHLPVVEL
 VGLHCKPKAPLPKMEGPFVOSGSDGVVVFISGVSNSEERAVVIALAKTIPOK
 VLMPREKRPDITGLNTQLYKMLPQNDLGLHPRTNAPFITHGNALIEALYHGIPIWVG
 VPLFADQNDIANKAKARVSLDPNMTSTDLHLAKTVINDPVENMKLSIHH
 DQVPLDRAVFWIEFVMRHRSKHLRVAAYDLTWQYHSLDIVGLFLACVAVIFII
 TKCLFVLKFEVFRGKKGRD"

BASE COUNT 564 a 374 c 404 g 573 t
 ORIGIN
 Query Match 76.0%; Score 1254.6; DB 9; Length 1915;
 Best local Similarity 87.0%; Pred. No. 3.6e-310;
 Matches 1390; Conservative 0; Mismatches 204; Indels 3; Gaps 1;
 QY 35 ATGCTTATGAATGAGCTTCTTCTCTGATACGCTGAGCTTACTTAACTCT 94
 DB 1 ATGCTTATGAATGAGCTTCTTCTCTGATACGCTGAGCTTACTTAACTCT 60
 QY 95 GGGAGTGTGGAAGGCTGCTGCTGAGCCCAAGAAATTCAGCACTGATGAATTAAG 154
 DB 61 GGGATTTGTGAAGGCTGCTGCTGAGCCCAAGAAATTCAGCACTGATGAATTAAG 120
 QY 155 ACAATCTGATGAACCTGTCCAGAGAGCTGATGAGGTGACTGATTTGAGCT 214
 DB 121 ACAATCTGATGAGCTGTCCAGAGAGCTGATGAGGTGACTGATTTGAGCT 180
 QY 215 TCCATTTCTTGCATCCCAACAGCCATCTTCTTAAATTTGAAGTTATCTGTATCT 274
 DB 181 TCCATTTCTTGCATCCCAACAGCCATCTTCTTAAATTTGAAGTTATCTGTATCT 240
 QY 275 TTAAGTAAAGTGAAGTTGAGATATTTATCAACAGCTGCTTAAAGATGGCAACTT 334
 DB 241 TTAAGTAAAGTGAAGTTGAGATATTTATCAACAGCTGCTTAAAGATGGCAACTT 300
 QY 335 CCAAAAGACACATTTTGGTCATATTTTTCACAGATCAAGAAATCATGTGACATTTAAT 394
 DB 301 CCAAAAGACACATTTTGGTCATATTTTTCACAGATCAAGAAATCATGTGACATTTAAT 360

```

Qy 395 GACATCTTAGAAGTCTGTAGAGATATAGTTTCAATTAAGAACTTATGAGAACTA 454
Db 361 GACATGATTAGAAAGTTCTGTAAAGATGTGTTTCTAATTAAGAACTTATGAGAACTA 420
Qy 455 CAGAGTCAGAAATTTGATGTGTCTTTCAGATGCTGTCTTCCCTTGTGTAGCTGTG 514
Db 421 CAGAGTCAGAAATTTGATGTGTCTTTCAGATGCTGTCTTCCCTTGTGTAGCTGTG 480
Qy 515 GCCGATTAATTAATATACCTTGTCTACAGCTTCCCTTGTCTGTCTGTAGCAATT 574
Db 481 GCTGAGCTAATTAATATACCTTGTGTACAGCTTCCCTTGTCTGTGTAGCAATT 540
Qy 575 GAAAGCATAGTGGAGGAGCTTCTGTTCCCTCTCTTCTATGTCGCTGTATGTCAGAA 634
Db 541 GAAAGCATAGTGGAGGAGTTCCTGTTCCCTCTCTCTATGTCGCTGTATGTCAGAA 600
Qy 635 CTAAGTCAGCAAAATGACTTTCATAGAGAGGTAATAATATGATCTATGTCTTATTTT 694
Db 601 TTCAGTATCAAAATGACTTTCATAGAGAGGTAATAATATGATCTATGTCTTATTTT 660
Qy 695 GAATTTTGTCTTCAAAATTTTTCATGAGAGAGTGGATCACTTCTACAGTGAAGTTCTA 754
Db 661 GACTTTTGTCTTCAAAATTTTTCATGAGAGAGTGGATCACTTCTACAGTGAAGTTCTA 720
Qy 755 GGAAGACCCACTAGCTTATCTGAGACATGCGAAAGCTGACATATGGCTTATGGAAC 814
Db 721 GGAAGACCCACTAGCTTATCTGAGACATGCGAAAGCTGACATATGGCTTATGGAAC 780
Qy 815 TACTGGATTTTTCATTTCTCTCACCCTCTTACCAAAATGTTGAGTGTGTGAGAGCTC 874
Db 781 TACTGGATTTTTCATTTCTCTCACCCTCTTACCAAAATGTTGAGTGTGTGAGAGCTC 840
Qy 875 CACTGCAAACTGCGCAAAACCCCTTACCGAAAGAGATGAAAGATTTGTCCAGAGCTTGA 934
Db 841 CACTGCAAACTGCGCAAAACCCCTTACCGAAAGAGATGAAAGATTTGTCCAGAGCTTGA 900
Qy 935 GAAATATGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 994
Db 901 GAAATATGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 860
Qy 995 GCCAATGTAATTTGATCAGACCTTCCCAAGATCCCAAAAGTTCTGTGAGATTTGAT 1054
Db 961 GCCAATGTAATTTGATCAGACCTTCCCAAGATCCCAAAAGTTCTGTGAGATTTGAT 1020
Qy 1055 GGGAAATTAACAGATCTTGAAGACTCAATATCGGCTGTACAGTGTGATCCCAAGAT 1114
Db 1021 GGGAAATTAACAGATCTTGAAGACTCAATATCGGCTGTACAGTGTGATCCCAAGAT 1080
Qy 1115 GATCTTCTTGTGATCCCAAAACCAAAAGCTTTTATCATCATGATGGAATGGAATG 1174
Db 1081 GATCTTCTTGTGATCCCAAAACCAAAAGCTTTTATCATCATGATGGAATGGAATG 1140
Qy 1175 TATGAAGTATTTTACATGAGGCTTCCCTATGTGTGAGTTCCTATTTGTGATCAGTT 1234
Db 1141 TATGAAGTATTTTACATGAGGCTTCCCTATGTGTGAGTTCCTATTTGTGATCAGTT 1200
Qy 1235 GATTAATAGTCTACATGAGGCTTCCCTATGTGTGAGTTCCTATTTGTGATCAGTT 1294
Db 1201 GATTAATAGTCTACATGAGGCTTCCCTATGTGTGAGTTCCTATTTGTGATCAGTT 1260
Qy 1295 ACAACGAAAGATTTTCTGAGGCTTGTGAGAGTCAATGATTCAGATTTCTTATTAAGG 1354
Db 1261 ACAACGAAAGATTTTCTGAGGCTTGTGAGAGTCAATGATTCAGATTTCTTATTAAGG 1320
Qy 1355 AATGCTATGAGTATCAAGATTTCAATGATTCATGATCAAGTGAAGCTTATGAGAGA 1414
Db 1321 AATGCTATGAGTATCAAGATTTCAATGATTCATGATCAAGTGAAGCTTATGAGAGA 1380
Qy 1415 GTCTTCTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1474
Db 1381 GTCTTCTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440

```

```

Qy 1475 CATGACCTCACTGTGTCAGACACTACTATATAGATGTATGGTTCCTGTCAGCTGT 1534
Db 1441 TAGACCTCACTGTGTCAGACACTACTATATAGATGTATGGTTCCTGTCAGCTGT 1500
Qy 1535 GTGGCAACTGCTATATTTCTTGTTCACAAATGTTTATTTTCTCTGTCAAAATTTAT 1594
Db 1501 GTGGCAACTGCTATATTTCTTGTTCACAAATGTTTATTTTCTCTGTCAAAATTTAT 1557
Qy 1595 AAAACTGAAAGATAGAAAAGAGGGAATAGATCTTTTC 1631
Db 1558 AGAACAGAGAAAGAGGAAAGAGATTTATATCATC 1594

RESULT 12
AX336329 1855 bp DNA linear PAT 09-JAN-2002
LOCUS AX336329
DEFINITION Sequence 6838 from Patent WO0194629.
ACCESSION AX336329
VERSION AX336329.1 GI:18127048
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 6838 13-DEC-2001;
JOURNAL Avalon Pharmaceuticals (US)
FEATURES
source
1. 1855
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 571 a 337 c 394 g 553 t
ORIGIN
Query Match 75.9%; Score 1252.6; DB 6; Length 1855;
Beet Local Similarity 86.1%; Pred. No. 1.2e-309;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
Qy 21 TGCAATGATAGAGATGCTATGAAATGAGCTTCTCTCTGATAGAGTGAAGT 80
Db 1 TGCAATGATAGAGATGCTATGAAATGAGCTTCTCTCTGATAGAGTGAAGT 60
Qy 81 GTTACTTTAGCTCTGGAGGTTGTGAAAGGTTGTGTGTGTGTGTGTGTGTGTGTGT 140
Db 61 TTTGCTTTAGCTCTGGAGGTTGTGAAAGGTTGTGTGTGTGTGTGTGTGTGTGTGT 120
Qy 141 GATGAATATTAAGACATCTCTGATGATGAACTTGTCCAGAGAGTATAGAGTGAAGT 200
Db 121 GATGAATATTAAGACATCTCTGATGATGAACTTGTCCAGAGAGTATAGAGTGAAGT 180
Qy 201 TGGCATCTTCAAGCTTCCATTTCTTTCATGCCAAGCCCACTCTTAAATTTGAAG 260
Db 181 TGGCATCTTCAAGCTTCCATTTCTTTCATGCCAAGCCCACTCTTAAATTTGAAG 240
Qy 261 TTTATCTGTATCTTTAACTTAACTGAGTGTGAGATTTATCAAGAGCTGTTAAGA 320
Db 241 TTTATCTGTATCTTTAACTTAACTGAGTGTGAGATTTATCAAGAGCTGTTAAGA 300
Qy 321 GATGGGCAAGCTTCCAAAGACATTTTGTGCTATTTTTCACAAAGTGAAGATCA 380
Db 301 GATGGGCAAGCTTCCAAAGACATTTTGTGCTATTTTTCACAAAGTGAAGATCA 360
Qy 381 TGTGCAATTTAATGACATCTTAAAGATCTGTAGAGATATAGTTCATTAAGAAAC 440
Db 361 TGTGCAATTTAATGACATCTTAAAGATCTGTAGAGATATAGTTCATTAAGAAAC 420
Qy 441 TTATGAGAAATCAAGAGTCAAGATTTGATGTGTCTTTCAGATGCTGTTTCCCT 500

```

```

Db 421 TTATGAAAAAGTACAGAGTCAAGATTGACGTATTTTTCAGAGTCAATTTTCCCT 480
Qy 501 TTGGTGAAGCTCTGGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTC 560
Db 481 GTAGTGAGCTCTGGCTGAGCTATTAAACATACCTTTGTGTACAGTCTCAGCTTCTC 540
Qy 561 CTGGCTACGCAATTGAAAAAGCATAGTGAGAGCTTCTGTTCCCTCTCTATGTGCTG 620
Db 541 CTGGCTACATTTTGAANAAGCATAGTGAGAGATTATTTTCCCTCTCTACGTACTG 600
Qy 621 TTGTTATGTCAAGTCAAGTCAACAAATGACTTTTCATAGAGAGGTAAAAATATGATCT 680
Db 601 TTGTTATGTCAAGTCAAGTCAACAAATGACTTTTCATAGAGAGGTAAAAATATGATCT 660
Qy 681 ATGCTTTATTTGAAATTTGTTGTTCCAAATTTGACATGAAGAAGGGATGATGTTCT 740
Db 661 ATGCTTTATTTGACTTTGACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 720
Qy 741 ACAGTGAAGTTCTAGAGAACCCACTACGTTATCTGAGACAAATGSCAAAGCTGACATAT 800
Db 721 ATAGTGAAGTTCTAGAGAACCCACTACGTTATCTGAGACAAATGSCAAAGCTGACATAT 780
Qy 801 GCGTTATTCGAATCTAGGAGTTTTCATTTCTCACCACCTTTACCAATGTTAAGT 860
Db 781 GCGTTATTCGAATCTAGGAGTTTTCATTTCTCACCACCTTTACCAATGTTAAGT 840
Qy 861 TCGTTGAGAGCTCAGTGCACAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTG 920
Db 841 TTGTTGAGAGCTCAGTGCACAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTG 900
Qy 921 TCCAGAGCTCTGAGAAAATGTTGTGTGTGTTTCTGAGGGTGCATGTTGCTAGTACA 980
Db 901 TAGAGAGCTCTGAGAAAATGTTGTGTGTGTTTCTGAGGGTGCATGTTGCTAGTACA 960
Qy 981 CGTCAAGAAAGGCGCAATGTAATGTCATAGCCCTTGCCAGATCCCAAAAAGTTTC 1040
Db 961 TGACAGAAAGGCGCAATGTAATGTCATAGCCCTTGCCAGATCCCAAAAAGTTTC 1020
Qy 1041 TGTCGAGTTTATGAGGGAATTAACCAATCTTATAGATCAATACGCTGCTGATACAGT 1100
Db 1021 TGTCGAGTTTATGAGGGAATTAACCAATCTTATAGATCAATACGCTGCTGATACAGT 1080
Qy 1101 GGATACCCCAAGATGATCTTCTGTGTCATCCCAAAAACCAAGCTTTTATCACTGATG 1160
Db 1081 GGATACCCCAAGATGATCTTCTGTGTCATCCCAAAAACCAAGCTTTTATCACTGATG 1140
Qy 1161 GAATGAATGGATGTATGAACTATTAACATGAGGCTCCATGCTGAGGAGTTCCATAT 1220
Db 1141 GAGCAATGGATGTATGAACTATTAACATGAGGCTCCATGCTGAGGAGTTCCATAT 1200
Qy 1221 TTGGTATCAGCTTGAATACATACCTCAATGAAGGCGCAAGAGAGCTGTAGAAATTA 1280
Db 1201 TTGGTATCAGCTTGAATACATACCTCAATGAAGGCGCAAGAGAGCTGTAGAAATTA 1180
Qy 1281 ACTTCAAAAAGTATGACAGAGATTTTACGAGGCTTTGAGAACAGTCTTACCGATT 1340
Db 1261 ACTTCAAAAAGTATGACAGAGATTTTACGAGGCTTTGAGAACAGTCTTACCGATT 1320
Qy 1341 CCTCTTAATAAGAGATGCTATGAGATTAATCAAGAAATTCACATGATCAACCTGTAAAG 1400
Db 1321 CCTCTTAATAAGAGATGCTATGAGATTAATCAAGAAATTCACATGATCAACCTGTAAAG 1380
Qy 1401 CCTCTTAATAAGAGATGCTATGAGATTAATCAAGAAATTCACATGATCAACCTGTAAAG 1460
Db 1381 CCTCTTAATAAGAGATGCTATGAGATTAATCAAGAAATTCACATGATCAACCTGTAAAG 1440
Qy 1461 TGCGATAGCTGCGCCATGACCTCACTGCTTCCAGAGCTACTATAGATGTAATGAGT 1520
Db 1441 TTGGGTTGAGAGCCACGACCTCACTGCTTCCAGAGCTACTATAGATGTAATGAGT 1500
Qy 1521 TCTGCTGAGCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCT 1580
Db 1501 TCTGCTGAGCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCT 1560

```

```

Qy 1581 GTCAAAAATTTTAAATTAATAAAGTATAGAAAAGAGGAGTATGATCTTC 1631
Db 1561 TCTGGAAGTTTGTCTAGAAAAGCAAAAGAGGAAAAATGATTTGATATATC 1611

RESULT 13
AX336696 1855 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 7205 from Patent WO0194629.
DEFINITION AX336696
ACCESSION AX336696
VERSION AX336696.1 GI:18127415
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 7205 13-DEC-2001;
Avalon Pharmaceuticals (US)
Location/Qualifiers
FEATURES
source
1..1855
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 571 a 337 c 394 g 553 t
ORIGIN

Query Match 75.9%; Score 1252.6; DB 6; Length 1855;
Best Local Similarity 86.1%; Pred. No. 1.2e-309;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 21 TGCAATGCATCAGAGATGTCTATGAAATGAGACTTCACTCTTCTCTGATACAGCTGAGT 80
Db 1 TGCAATGCATCAGAGATGTCTATGAAATGAGACTTCACTCTTCTCTGATACAGCTGAGT 60
Qy 81 GTTACTTTAGCTCTGGAGTGTGGAAGGTGCTGTGTGGCCCAAGAAATTCAGCCACT 140
Db 61 TTGCTTTAGCTCTGGAGATTTGGAAGGTGCTGTGTGGCCCAAGAAATTCAGCCACT 120
Qy 141 GGAATGAATTAAGACAAATCTCGATGAACTTGTCCAGAGAGGTCATGAGAGTATAT 200
Db 121 GGAATGAATTAAGACAAATCTCGATGAACTTGTTCAGAGAGGTCATGAGAGTATAT 180
Qy 201 TGGCATCTTCAGCTTCATTTCTTTGATCCCAAGAGCCCATCTACTTAAATTTGAG 260
Db 181 TGGCATCTTCAGCTTCATTTCTTTGATCCCAAGAGCCCATCTACTTAAATTTGAG 240
Qy 261 TTATCCCTGATCTTTAACTAAAGTGAAGTTTGAAGATATTAATCAAGAGCTGTTAGA 320
Db 241 TTATCCCTGATCTTTAACTAAAGTGAAGTTTGAAGATATTAATCAAGAGCTGTTAGA 300
Qy 321 GATGGGAGAACTTCCAAAAGACATTTTGTGCTATTTTTCAGAAATCAAGAAATCA 380
Db 301 GATGGGAGAACTTCCAAAAGATCAATTTTGTGCTATTTTTCAGAAATCAAGAAATCA 360
Qy 381 TGTCGACATTTATGACATCTTGAAGAGTTCTGTAAGATATAGTTTCAATTAAGAAC 440
Db 361 TGTCGACATTTATGACATCTTGAAGAGTTCTGTAAGATATAGTTTCAATTAAGAAC 420
Qy 441 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTTCAGAGTGTGTTCCCT 500
Db 421 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTTCAGAGTGTGTTCCCT 480
Qy 501 TTGGTGAAGCTGTGGCCGAGTTACTTAAATTAACCTTTGTCTACAGCTCCGCTCTC 560
Db 481 GTAGTGAAGCTGTGGCCGAGTTACTTAAATTAACCTTTGTCTACAGCTCCGCTCTC 540
Qy 561 CTGGCTACGCAATTGAAAAAGCATAGTGAGAGCTTCTGTTCCCTCTCTATGTGCTG 620

```

```

Db      ||||| 541 CTGGCTACACTTTGAAAAGCATAGTGGAGATTATTTCCCTCCTTACGACGCTG 600
Qy      ||||| 621 TTGTTATGTCGAACTAGTACCCAAATGACTTTTCATGAGAGGGTAAATAATGATCT 680
Db      ||||| 601 TTGTTATGTCGAAATTAATGATCAATGACTTTTCATGAGAGGGTAAATAATGATCT 660
Qy      ||||| 681 ATGCTCTTATTTGAAATTTGGTTCGAAATATTTGATGAGAGAGGGATCAGTCT 740
Db      ||||| 661 ATGCTCTTATTTGAACTTTGGTTCGAAATATTTGATGAGAGAGGGATCAGTCT 720
Qy      ||||| 741 ACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATAT 800
Db      ||||| 721 ATAGGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATAT 780
Qy      ||||| 801 GCGTTATTCGAAATCTAGGAGTTTTCATTTCTCCACCCACTCTTACCAATGTTGAGT 860
Db      ||||| 781 GCGTTATTCGAAATCTAGGAGTTTTCATTTCTCCACCCACTCTTACCAATGTTGAGT 840
Qy      ||||| 861 TCGTTGAGAGACTCCACTGCAAACTGGCAAACTCCGACGAAAGGAATGGAAGAGTTG 920
Db      ||||| 841 TTGTTGAGAGACTCCACTGCAAACTGGCAAACTCCGACGAAAGGAATGGAAGAGTTG 900
Qy      ||||| 921 TCCAGAGCTCTGAGAAAATGGTGTGTGTTGTTCTCTGGAGTCTGATGCTGAGTAA 980
Db      ||||| 901 TACAGAGCTCTGAGAAAATGGTGTGTGTTGTTCTCTGGAGTCTGATGCTGAGTAA 960
Qy      ||||| 981 CGTCAAGAAAAGGCGCAATGTAATTCATGACAGCTTTCGCAAGATCCCAAAAAGTTTC 1040
Db      ||||| 961 TGACGAAAGAAAAGGCGCAATGTAATTCATGACAGCTTTCGCAAGATCCCAAAAAGTTTC 1020
Qy      ||||| 1041 TGTGAGATTTGATGGGAATTAACAGATTAATTTGAGACTCAATGCTGAGTCTGATCAAGT 1100
Db      ||||| 1021 TGTGAGATTTGATGGGAATTAACAGATTAATTTGAGACTCAATGCTGAGTCTGATCAAGT 1080
Qy      ||||| 1101 GGATACCCGAGAAATGATCTTTTGGTCAATCCCAAAACCAAAAGCTTTTACTGCTGAGT 1160
Db      ||||| 1081 GGATACCCGAGAAATGATCTTTTGGTCAATCCCAAAACCAAAAGCTTTTACTGCTGAGT 1140
Qy      ||||| 1161 GAATGAATGGGATCTATGAGATTTTACATGAGGGGCTCCATGATGGTGGAGTTCCCATAT 1220
Db      ||||| 1141 GAGCCAAATGGGATCTATGAGATTTTACATGAGGGGCTCCATGATGGTGGAGTTCCCATAT 1200
Qy      ||||| 1221 TTGGATCAGCTGATTAACATAGCTCAATGAGAGGCAAGAGAGAGAGAGTGAATTA 1280
Db      ||||| 1201 TTGGATCAGCTGATTAACATAGCTCAATGAGAGGCAAGAGAGAGAGAGTGAATTA 1260
Qy      ||||| 1281 ACTTCAAAATCTATGACAAAGCAAGATTTACTGAGGGCTTTGAGAAAGTCAATTAACGAT 1340
Db      ||||| 1261 ACTTCAAAATCTATGACAAAGCAAGATTTACTGAGGGCTTTGAGAAAGTCAATTAACGAT 1320
Qy      ||||| 1341 CCTCTTAAAGAGAAATGCTATGAGATTTCAAGAAATTCACATGATCAACCTGTAAGC 1400
Db      ||||| 1321 CCTCTTAAAGAGAAATGCTATGAGATTTCAAGAAATTCACATGATCAACCTGTAAGC 1380
Qy      ||||| 1401 CCTGATGAGAGAGCTTTCTGAGATGAGTTGTGATGGGCAAAAGAGGCAAGACACC 1460
Db      ||||| 1381 CCTGATGAGAGAGCTTTCTGAGATGAGTTGTGATGGGCAAAAGAGGCTTAACACC 1440
Qy      ||||| 1461 TCGCATGAGAGCTCCCATGACCTCACTGCTGTTCCAGACACTCTATAGATGATGGGT 1520
Db      ||||| 1441 TTCCGGGTTGCAAGCCCAAGACCTCACTGCTGTTCCAGACACTCTATAGATGATGGGT 1500
Qy      ||||| 1521 TCTCTGCTGACCTGTGTGGCACTGCTATATCTTGTTCACAAAATGTTTATTTTCT 1580
Db      ||||| 1501 TCTCTGCTGCTGTGTGGCACTGCTATATCTTGTTCACAAAATGTTTATTTTCT 1560
Qy      ||||| 1581 GTCAAAAATTTTATAAATCTAGAAGATGAAAAGGGAATAGATCTTTC 1631
Db      ||||| 1561 TCTGGAAGTTGCTAGAAAAGCAAGAGGGAATAGATTAATATTC 1611

```

```

AX409473
LOCUS      AX409473              1855 bp      DNA              linear      PAT 14-JUN-2002
DEFINITION Sequence 2120 from Patent WO0229103.
ACCESSION  AX409473
VERSION     AX409473.1
KEYWORDS    GI:21442178
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
            Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
            Gene expression profiles in liver cancer
            Patent: WO 0229103-A 2120 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES
    source
        location/Qualifiers
            1..1855
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /note="EMBL/GenBank Accession No. J05428"
BASE COUNT  571 a      337 c      394 g      553 t
ORIGIN
Query Match      75.9%; Score 1252.6; DB 6; Length 1855;
Best Local Similarity 86.1%; Pred. No. 1,26-309;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
21  TGCATTGCATCAGATGCTCTATGAAATGAGCTTACGCTCTTCTCTGATACAGTGAAGT 80
Db  1  TGCATTGCATCAGATGCTCTGTAATGAGCTTCAAGTAATTTTCTAATACACTGAGCT 60
Qy  81  GTTACTTTAGCTCTGGGAGTTGTGAAAAGGCTGTGTGTGGCCGACAGAAATTCAGCACT 140
Db  61  TTTGCTTTAGCTCTGGGAGTTGTGAAAAGGCTGTGTGTGGCCGACAGAAATTCAGCACT 120
Qy  141  GGATGAATATTAAGACAATCTGAGTGAATCTGTCAGAGAGTCTATGAGTGAATCTGAT 200
Db  121  GGATGAATATTAAGACAATCTGAGTGAATCTGTCAGAGAGTCTATGAGTGAATCTGAT 180
Qy  201  TGGCATCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCTCTACTCTTAAATTTGAAG 260
Db  181  TGGCATCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCTCTACTCTTAAATTTGAAG 240
Qy  261  TTTATCCGTATCTTTAACTAAACCTGAGTTGAGATTAATATCAAGAGCTGGTTAAGA 320
Db  241  TTTATCCGTATCTTTAACTAAACCTGAGTTGAGATTAATATCAAGAGCTGGTTAAGA 300
Qy  321  GATGGGCAAGACTTCCAAAAGACATTTTGTGATATTTTTCACAAAGTACAGAAATCA 380
Db  301  GATGGGCAAGACTTCCAAAAGACATTTTGTGATATTTTTCACAAAGTACAGAAATCA 360
Qy  381  TGTGACATTTAATGACATCTTAGAAAGTCTGTAAAGATATAGTTTCAATTAAGAAAC 440
Db  361  TGTGACATTTAATGACATCTTAGAAAGTCTGTAAAGATATAGTTTCAATTAAGAAAT 420
Qy  441  TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTCCCT 500
Db  421  TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTCCCT 480
Qy  501  TTGTGAGAGCTGCTGGCCAGTTACTTAATAATCCCTTGTCTAGAGCTCGCTTCTCTC 560
Db  481  GTAGAGAGCTGCTGGCCAGTTACTTAATAATCCCTTGTCTAGAGCTCGCTTCTCTC 540
Qy  561  CTGGCTACGCAATTGAAAAGCATAGTGAAGACTTGTCTCTCTCTTCTCTATGCTGCTG 620
Db  541  CTGGCTACGCAATTGAAAAGCATAGTGAAGACTTGTCTCTCTCTCTCTATGCTGCTG 600
Qy  621  TTGTTATGTCGAGACTAGTACCAATGACTTTTCATGAGAGGGTAAATAATGATCT 680
Db  601  TTGTTATGTCGAGACTAGTACCAATGACTTTTCATGAGAGGGTAAATAATGATCT 660
Qy  681  ATGCTCTTATTTGAAATTTGGTTCCAAAATATTTGACATGAAGAAAGTGGATCAGTTCT 740

```


Dp	661	ATGTCCTTACTTTGACTTTGGTTGCGAAATTTTGGACATGAAGAAGTGGATCAGTTT	720
Oy	741	ACAGTGAAGTCTAGAAAGACCCTACGTATCTGAGCAATGGCCAAAGCTGACATAT	800
Dp	721	ATAGGAAGTCTAGAAAGACCCTACGTATCTGAGCAATGGGGAAGCTGACGAT	780
Oy	801	GGCTTATTCGAAACTACTGGGATTTTCATTTTCTCACCCTTTTACCAATGTTGAGT	860
Dp	781	GGCTTATTCGAAACTCTGGAATTTTCAGTTTCTCATTCACCTTTACCAATGTTGAT	840
Oy	861	TCGTTGAGGAGCTCCACTGGCAACCTGGCAAAACCCCTACCGAAGGAATGGAAGTTTG	920
Dp	841	TTGTTGGAGGACTCCACTGGCAACCTGGCAAAACCCCTGCTTAAAGGAATGGAAGCTTGG	900
Oy	921	TCGAGAGCTCTGGAGAAAATGSGTGTGGTGATTTTTCCTGGGGTGCATGCGCACTAAC	980
Dp	901	TACAGAGCTCTGGAGAAAATGSGTGTGGTGATTTTTCCTGGGGTGCATGCGCACTAAC	960
Oy	981	CGTCGAGAGAAAGGGCCCAATGTAATGTCATCAGCCCTTGCAGATGCCCAAAAGTTTC	1040
Dp	961	TGACGAGAGAAAGGGCCCAAGTAATGTCATCAGCCCTGCGCAATGCCCAAAAGTTTC	1020
Oy	1041	TGTGAGATTTGATGGGAATTAACGAGATCTTTTGGACTCAATCTGGCTGTCAAGT	1100
Dp	1021	TGTGAGATTTGATGGGAATTAACGAGATCCTTTGGCTCAATCTGGCTGTCAAGT	1080
Oy	1101	GGATGACCCAGATATCTCTGGTCAATCCCAAAACCAAGCTTTTATCACTCATGSGT	1160
Dp	1081	GGATGACCCAGATATCCTCTGAGTCAATCCCAAAACCAAGACTTTTATCACTCATGSGT	1140
Oy	1161	GAATGAATGGGATCTATGAGCTATTTACATGGGGTCCCTATGSGTGGAGTTCCCATAT	1220
Dp	1141	GAGCCAAATGGCATCTACGAGGCAATCTACATGGGATCCCTATGSGTGGAGTTCCCAT	1200
Oy	1221	TTGTGATCAGCTTGATTAACATAGTCAACATGAAGGCCAAAGAGCAGCTGTGAATA	1280
Dp	1201	TTGCGGATCAACCTGATTAACATGTCTCAATGAAGGCCAGGGGACAGCTGTGAAGT	1260
Oy	1281	ACTTCAAAACTATGACCAAGCAAGTTTACTGAGGGCTTTGAAACAGTCAATACCAT	1340
Dp	1261	ACTTCAACACATGTGCAATGCAAGCTTGTCTGAATGCAATGGAAGAGATTAATGATC	1320
Oy	1341	CCTCTTATTAAGAGATGCTATGATTAATCAAGATTAACCATATCAACCTGTAAGC	1400
Dp	1321	CTTCAATTAAGAGATGCTATGATTAATCAAGATTAACCATATCAACAGTAAGC	1380
Oy	1401	CCCTAGATCGAGAGCTCTTCTGGAATGCTGATCGGCACAAAGAGCAGACACC	1460
Dp	1381	CCCTGAGATCGAGAGCTCTTCTGGAATGCTGATCGGCACAAAGAGCTTAACACC	1440
Oy	1461	TGCGATCAGTCCGCAATGACCTCACCTGGTTCCAGCACTACTCTATPAGATGATGGGT	1520
Dp	1441	TTGCGGTTGACAGCCACGACCTCACCTGGTTCCAGTACACGCTTTTGATGATGGGT	1500
Oy	1521	TCCTGCTAAGCTGTGTGGCACTGCTATATTTCTGTCACAAAATGTTTTTATTTCT	1580
Dp	1501	TCCTGCTGTGTGTGTGGCACTGCTATATTTCTGTCACAAAATGTTGTCTGTTTGT	1560
Oy	1581	GTCAAAAATTTAATTAACCTGAAGATGAGAAAAGAGGATAGATCTTTC	1641
Dp	1561	TCCTGGAATTTGCTAGAAAAGCAAAAGGGAATAATGATTAATATC	1611

RESULT 15	HUMUDPCTA	1855 bp	mRNA	linear	PRI 03-AUG-1993
LOCUS	Human 3,4-catechol estrogen UDP-glucuronosyltransferase mRNA,				
DEFINITION	complete cds.				
ACCESSION	J05428				
VERSION	J05428.1	GI:340079			
KEYWORDS	3,4-catechol estrogen UDP-glucuronosyltransferase.				
SOURCE	Homo sapiens (human)				

ORANISM	Homio sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1855)
TITLE	Ritter, J.K., Sheen, Y.Y. and Owens, I.S. Cloning and expression of human liver UDP-glucuronosyltransferase in COS-1 cells. 3,4-catechol estrogens and estril as primary substrates
JOURNAL	J. Biol. Chem. 265 (14), 7900-7906 (1990)
MEDLINE	90243659
PUBMED	2159463
COMMENT	Original source text: Human liver, cDNA to mRNA, clone 63-11. Draft entry and computer-readable sequence for [1] kindly submitted by I.S.Owens, 22-FEB-1990.
FEATURES	Location/Qualifiers
source	1..1855
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	15..1604
CDS	/note="UDP-glucuronosyltransferase (EC 2.4.1.17)"
	/codon_start=1
	/protein_id="AA36793.1"
	/db_xref="GI:340080"
	/translation="MSYKTSVLLIQLSPFSFGNGKVLVMAEYSHMNIKTILD ELIQRHEVTLVASSASILLPDPNNSSALKEIETSLTKLELNFTIMQIKRSDLPK DTIMLYRSQVQIEMISIFGDIITRKPCRDVSNKKPFMKVQSRDVI PADAIPOSELL AELFNPIEYVLSFSFGYTEKXSGGFIPEPSVPVMSLTQDMFEMVKMIYVL YFPEFWEIPDMKMDQFESYEVIGRPTLLSTMKADWMLIRSNWNRQFPBHLPNDF VGGHCKEAPKLEKMEDEFGSENGVVFSGMSGNSTERRAVIYASALQIPQK VLRFPGNKPDGTGLNTRILKYMI PONDLDHPTPRAFI TGGAANGYIEAYHYGIPWG IPLPADQNPILAHMKRAGVAVRDPNMTSTDLNLKRYINDPSKENMTKSLRQH DQKPKLDRVFNIEFVMRKKGAKHLRVAAHDLTWQYHSLDIVGLVCVAATVIFIV TKCCLEPFMKFAKKKGNKD"
BASE COUNT	571 a 337 c 394 g 553 t
ORIGIN	
Query Match	75.9%; Score 1252.6; DB 9; Length 1855;
Best Local Similarity	86.1%; Pred. No. 1.2e-309;
Matches 1387; Conservative	0; Mismatches 224; Indels 0; Gaps 0;
Ds	21 TGCATTGCATCAGAGTGTCTATGAAAGAGACTTCAGCTCTTCTCTGATACAGGTGAGCT 80
Ds	1 TGCATTGCACAGAGATCTGTGAAATGAGACTTCAGTAATTTTGTCTAATCAACTGAGCT 60
Ds	81 GTTACTTACTGCTGGAGATTGGAAAGGTGCTGGTGGCCCAAGAAATTACGCACT 140
Ds	61 TTGGCTTTTACTCTGGGAATTGTGGAAAGGTGCTGGTGGCGAGCAGAAATTAAGCCATT 120
Ds	141 GGATGAATTAAAGACAACTCTGGATGAATCTTGCCAGAGAGGTCATGAGTACGTGTAT 200
Ds	121 GGATGAATTAAAGACAACTCTGGATGAATCTTATTCAGAGAGTCAATGAGGTGACGTGAC 180
Ds	201 TGGCATCTTCAGCTTCCTCAATTTCTTGGATCCCAACAGCCCATCTACTCTTAAATTTGAAG 260
Ds	181 TGGCATCTTCAGCTTCCTCAATTTCTTGGATCCCAACAGCTATCGGCTTTAAATTTGAAA 240
Ds	261 TTTATCTCTGATCTTTTAACTAAAGATGATTTGAGATATTTATCAAGCAGCTGTTAAGA 320
Ds	241 TTTATCTCACATCTTTTAACTAAAGATGATTTGAGAAATTTTCATCATGCAACAATTAAAGA 300
Ds	321 GATGGGAGAACTTCCAAAAGACATTTTGGTCATATTTTTCACAAGTACAGAATAATCA 380
Ds	301 GATGGTACAGCTTCCAAAAGATGATTTTGGTATATTTTTCACAAGTACAGAATAATCA 360
Ds	361 TGTCAATATTTGGTACATTAATCAAGAACTTCTGTAAGATGTAGTTCAAAATPAAGAAAT 420
Ds	441 TTATGAAGAACTTAACAGATCAAGATTTGATGTGTGTTCTTGAGAGATGCTTTTCCCT 500
Ds	421 TTATGAAGAACTTAACAGATCAAGATTTGATGTGTGTTCTTGAGAGATGCTTTTCCCT 480

QY 501 TTGGAGACCTGCTGGCCGAGTTACTTAATAATACCTTTGTCTACAGACCTCCGCTCTCTC 560
DB 481 GTAGTGAAGCTGCTGCTGAGCTATTTAATACATACCTTTGTGTGACGTCTCAGCTTCTCTC 540
QY 561 CTGGCTACGCAATTTGAAAAGCATAGTGGAGAGACTTCTGTCTCTCTCTCTCTCTGCTG 620
DB 541 CTGGCTACGCAATTTGAAAAGCATAGTGGAGAGACTTCTGTCTCTCTCTCTCTCTGCTG 600
QY 621 TTGTTATGCTGCAATAGTACCAATGACCTTTCATGAGAGGTTAAATAATATGATCT 680
DB 601 TTGTTATGCTGCAATAGTACCAATGACCTTTCATGAGAGGTTAAATAATATGATCT 660
QY 681 ATGTGCTTATTTTGAATTTTGTCTCAAAATTTTGCATGAGAGAGTGGATCAGTTCT 740
DB 661 ATGTGCTTATCTTTGACTTTTGTCTCAAAATTTTGCATGAGAGAGTGGATCAGTTCT 720
QY 741 ACAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGCAATGAGCAAAAGCTGACATAT 800
DB 721 ATAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGCAATGAGGAAAGCTGACATAT 780
QY 801 GGCTTATTCGAATCACTAGGAGTTTCAATTTCTCAACCCACTTACCAATGTTGAGT 860
DB 781 GGCTTATTCGAATCACTAGGAGTTTCAATTTCTCAATCACTTACCAATGTTGAGT 840
QY 861 TCGTTGAGAGACTCCACTGCAAAACCTGCAAAACCCCTACCCGAGAGAAATGGAAGATTG 920
DB 841 TTGTTGAGAGACTCCACTGCAAAACCTGCAAAACCCCTGCTAAGAAATGGAAGACTTG 900
QY 921 TCCAGAGCTCTGAGAAAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 980
DB 901 TACAGAGCTCTGAGAAAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960
QY 981 CCTGAGAAAAGGAGCCATGTAATTTGCATGAGCCCTTGCAGAGATCCCAAAAGGTTT 1040
DB 961 TGACGAGAAAAGGAGCCATGTAATTTGCATGAGCCCTTGCAGAGATCCCAAAAGGTTT 1020
QY 1041 TGTGAGATTTGATGGAATTAACAGATTACTTAGACTCAATCTGAGTGTACAGT 1100
DB 1021 TGTGAGATTTGATGGAATTAACAGATTACTTAGACTCAATCTGAGTGTACAGT 1080
QY 1101 GGATACCCGAGATGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1160
DB 1081 GGATACCCGAGATGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1161 GAATGAATGGATCTATGAGATTTTACATGGGGTCCCTATGTTGGAGTTCCCATAT 1220
DB 1141 GAGCCAAATGGATCTATGAGATTTTACATGGGGTCCCTATGTTGGAGTTCCCATAT 1200
QY 1221 TTGATGATCAGCTTATTAACATAGCTCAATGAGGAGCCAAAGAGCAGCTGTAGAAATTA 1280
DB 1201 TTGATGATCAGCTTATTAACATAGCTCAATGAGGAGCCAAAGAGCAGCTGTAGAGTGG 1260
QY 1281 ACTTCAAACTATGACAAAGCAAGTTTACTGAGGGCTTTGAGAAAGCTATTAACGAT 1340
DB 1261 ACTTCAAACTATGACAAAGCAAGTTTACTGAGGGCTTTGAGAAAGCTATTAATGATC 1320
QY 1341 CCTCTTATTAAGAGATGCTATGAGATTTTCAAGAAATTTACATGATCAACCTGTAAAGC 1400
DB 1321 CCTCTTATTAAGAGATGCTATGAGATTTTCAAGAAATTTACATGATCAACCTGTAAAGC 1380
QY 1401 CCTGATGAGAGAGCTTCTGATGAGTTGTCAATGGGCCCAAAAGGAGCCAAAGCACC 1460
DB 1381 CCTGATGAGAGAGCTTCTGATGAGTTGTCAATGGGCCCAAAAGGAGCCAAAGCACC 1440
QY 1461 TGGGATCAGCTGCCATGACCTCACTGTTCCAGACACTACTTATAGATGATGGGT 1520
DB 1441 TGGGATGAGAGAGCCCAAGCAGCTCACTGTTCCAGACACTCTTGTGATGATGGGT 1500
QY 1521 TCTGCTGACCTGTGTGGCACTGCTATATTTCTGTTCAGAAAATGTTTATTTTCT 1580
DB 1501 TCTGCTGCTGTGTGGCACTGCTATATTTTATGCTCAAAAATGTTGTCTGTGTGT 1560
QY 1581 GTCAAAAATTTAATAAAGATAGAAAAGATGAAAAGAGGATAGATCTTTC 1631

DB 1561 TCTGAGATTGCTAGAAAAGCAAAAGGAAAATGATTAATATATC 1611

Search completed: December 7, 2003, 06:56:33
Job time : 4110 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 03:51:27 ; Search time 349 Seconds
(without alignments)
12762.366 Million cell updates/sec

Title: US-09-980-729B-10

Perfect score: 1650
Sequence: 1 agcacctgcgaataacacgacat.....ccaattccaagaagacctg 1650

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1650	100.0	21	AAC65396	Human carbohydrate
2	1445.4	87.6	2092	AAZ95199	Human UDP-glucuron
3	1437.6	86.8	2093	ABN97347	Gene #3945 used to
4	1428	86.5	2111	AA569712	DNA encoding novel
5	1331.4	80.7	1639	AAAL4190	Drug metabolising
6	1272.6	77.1	1991	AAAD4591	Human UGT2B7 cDNA.
7	1252.6	75.9	1855	ABN95622	Gene #2120 used to
8	1252.6	75.9	1855	ABL68501	Kidney cancer rela

9	1252.6	75.9	1855	24	ABL68868
10	1247.8	75.6	1854	21	AAZ95200
11	1238	75.0	1722	22	AAZ06860
12	1235.2	74.3	2739	24	ABN97261
13	1221.4	74.0	1713	22	AAZ06844
14	1207	73.2	1859	23	AA569710
15	1199.4	72.7	2802	23	AA569711
16	1159.6	70.3	2107	19	AAV15900
17	1126.8	68.3	1976	16	AAZ95206
18	1125.2	68.2	2090	24	ABK84210
19	1125.2	68.2	2090	24	ABN96795
20	1125.2	68.2	2090	25	ABT17080
21	1115	67.6	2123	24	ABN96789
22	967.4	58.6	1636	24	AAZ24666
23	967.4	58.6	1705	25	ABX72195
24	967.4	58.6	2759	24	ABK12422
25	967.4	58.6	2974	22	AA546185
26	967.4	58.6	2974	25	ACA57943
27	967.4	58.6	2974	25	ABX98413
28	967.4	58.6	2974	25	ABX98915
29	967.4	58.6	2974	25	ACA05960
30	967.4	58.6	2974	25	ABX98004
31	967.4	58.6	2974	25	ABX78788
32	967.4	58.6	2974	25	ABX75801
33	967.4	58.6	2974	25	ABX77006
34	967.4	58.6	2974	25	ABX16846
35	965.8	58.5	2966	25	ABZ58826
36	964.2	58.4	3050	25	ABX72194
37	956.8	58.0	1961	24	ABK63496
38	956.8	56.8	1584	24	ABZ58827
39	872.2	52.9	1606	24	ABN85391
40	836.8	50.7	1024	25	ABZ83776
41	829.2	50.3	1766	23	AA591380
42	818	49.6	1413	23	AAZ93928
43	754.8	45.7	1323	21	AAZ95193
44	610.8	37.0	1686	21	AAZ95201
45	572.4	34.7	1001	21	AAH51512

ALIGNMENTS

RESULT 1	
AAC65396	
ID AAC65396 standard; cDNA; 1650 BP.	
XX	
AC AAC65396;	
XX	
DT 13-FEB-2001 (first entry)	
XX	
DE Human carbohydrate-modifying enzyme cDNA Incyte ID No: 2912330CB1.	
XX	
XX Human, carbohydrate-modifying enzyme; CME; antidiabetic;	
KW immunosuppressive; anti-HIV; antiinflammatory; antiaiemic;	
KW antiaethmatic; antiarteriosclerotic; antithyroid; hepatotropic;	
KW nephrotropic; antipour; thymimetic; neuroprotective; osteopahic;	
KW antiaethmatic; antipouric; utopahic; ophtalmologic;	
KW dermatologic; antilucer; cytosatic; vitucide; antibacterial;	
KW fungicide; protozoacide; tranquilliser; vulnerary; diabetes;	
KW autoimmune disorder; inflammatory disorder; infection; ss.	
XX	
OS Homo sapiens.	
XX	
PN WO200063351-A2.	
XX	
PD 26-OCT-2000.	
XX	
PF 20-APR-2000; 2000WO-US10882.	
XX	
PR 21-APR-1999; 99US-0130383.	
XX	
PA (INCY-) INCYTE GENOMICS INC.	
XX	

Kidney cancer rela
Human UDP-glucuron
Human breast cancer
Gene #3759 used to
Human breast cance
DNA encoding novel
Uridine diphospho-
Human UDP-glucuron
Human cDNA differe
Gene #3293 used to
Androgen-independe
Gene #3287 used to
Human drug metabol
Human NOVX polynuc
cDNA encoding huma
Human DNA encoding
Human PRO6239 cDNA
Human cDNA encodin
Novel human secret
Human secreted/tra
Human PRO polynuc
Human PRO polynuc
Human cDNA encodin
Human PRO polynuc
Human 32624 polype
Human NOVX polynuc
Rat sequence diffe
Human NOV13, glucu
Toxicologically re
DNA encoding novel
Human UDP-glucosyl
Human UGT2B4 exon
Human UGT2B7 exon
Human UGT2B10 rela

PI Lai P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
 XX MPI: 2000-672729/65.
 DR P-PSDB: AAB28677.
 XX
 PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides
 PT for diagnosis, treatment, and prevention of carbohydrate metabolism
 PT disorders, autoimmune/inflammatory disorders, and cancer -
 XX
 SS Claim 4; Page 75; 75pp; English.
 XX
 CC The present cDNA sequence encodes a human carbohydrate-modifying enzyme
 CC (CME). CME polynucleotides and polypeptides are useful for treating and
 CC diagnosing diseases associated with CME such as diabetes,
 CC autoimmune/inflammatory disorders such as AIDS, Addison's disease,
 CC adult respiratory distress syndrome, allergies, anaemia, asthma,
 CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,
 CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
 CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
 CC Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic
 CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
 CC complications of cancer, haemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
 CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
 CC useful for drug screening.
 XX

Sequence 1650 BP; 489 A; 330 C; 354 G; 477 T; 0 other;

Query Match 100.0%; Score 1650; DB 21; Length 1650;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1650; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAACTGGAAACAGCACTTCATTCATCGATGCTGTAATGAGCTTCAGCTCT 60
 DB 1 AGCAACTGGAAACAGCACTTCATTCATCGATGCTGTAATGAGCTTCAGCTCT 60
 QY 1 TCTCTGATACAGCTGAGCTGTTACTTAACTCTGGAGTGTGGAAGGCTGGGTG 120
 DB 1 TCTCTGATACAGCTGAGCTGTTACTTAACTCTGGAGTGTGGAAGGCTGGGTG 120
 QY 121 GCCCAGCAATTCAGCCACTGGATGAATATAAGCAATCCGATGAATCTGCCAG 180
 DB 121 GCCCAGCAATTCAGCCACTGGATGAATATAAGCAATCCGATGAATCTGCCAG 180
 QY 121 AGGTCATGAGGTGACTGATTTGGCATCTTCAGCTTCATTTCTTCGATCCCAAGCC 240
 DB 121 AGGTCATGAGGTGACTGATTTGGCATCTTCAGCTTCATTTCTTCGATCCCAAGCC 240
 QY 241 ATCTACTCTTAATTTGAAGTTATCCGTAATCTTAATAACTGAGTTGAGGATAT 300
 DB 241 ATCTACTCTTAATTTGAAGTTATCCGTAATCTTAATAACTGAGTTGAGGATAT 300
 QY 301 TATCAGAGAGCTGTTAAGAGATGGGAGAACTTCCAAAAGACATTTGGTCAATTT 360
 DB 301 TATCAGAGAGCTGTTAAGAGATGGGAGAACTTCCAAAAGACATTTGGTCAATTT 360
 QY 361 TTCACAGTACAAAGATCATGTGACATTTAATGACATCTTAAGAAAGTTCTGTAAGA 420
 DB 361 TTCACAGTACAAAGATCATGTGACATTTAATGACATCTTAAGAAAGTTCTGTAAGA 420
 QY 421 TATATCTTCAATAAATAAGAACTTATAGAAACTAAGAGATCAAAATTTGATGTTCT 480
 DB 421 TATATCTTCAATAAATAAGAACTTATAGAAACTAAGAGATCAAAATTTGATGTTCT 480
 QY 481 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGCCAGATTAATTAATCCCTTTGT 540
 DB 481 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGCCAGATTAATTAATCCCTTTGT 540
 QY 541 CTACAGCCTCGCTTCTCTCTGCTGCTACGCAATTAAGAAAGATGAGAGACTTCTGT 600
 DB 541 CTACAGCCTCGCTTCTCTCTGCTGCTACGCAATTAAGAAAGATGAGAGACTTCTGT 600

QY 601 CCTCTTCCATATGCGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTCAATGA 660
 DB 601 CCTCTTCCATATGCGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTCAATGA 660
 QY 661 GAGGTAATAAATATGATCTATGCTGTTATTTTGAATTTGGTCCAAATATTTGACAT 720
 DB 661 GAGGTAATAAATATGATCTATGCTGTTATTTTGAATTTGGTCCAAATATTTGACAT 720
 QY 721 GAAGAATGGGATATGATTTTACAGTGAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 780
 DB 721 GAAGAATGGGATATGATTTTACAGTGAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 780
 QY 781 AATGGCAAAAGCTGACATATGCTTATTCGAACTACCTGGGATTTTCAATTTCTCACCC 840
 DB 781 AATGGCAAAAGCTGACATATGCTTATTCGAACTACCTGGGATTTTCAATTTCTCACCC 840
 QY 841 ACTCTTACCAATATGTTAGTTGCTGTTGAGAGACTTCACCTGCAAACTGCCAACCCCTAAC 900
 DB 841 ACTCTTACCAATATGTTAGTTGCTGTTGAGAGACTTCACCTGCAAACTGCCAACCCCTAAC 900
 QY 901 GAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAGAAATGTTGTTGTTTCTCT 960
 DB 901 GAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAGAAATGTTGTTGTTTCTCT 960
 QY 961 GGGGTGATGCTGCTGTAACAGCTCAGAAAGAGGCCAATGTAATGCAATGAGCCCTTGC 1020
 DB 961 GGGGTGATGCTGCTGTAACAGCTCAGAAAGAGGCCAATGTAATGCAATGAGCCCTTGC 1020
 QY 1021 CAAGATCCCAAAAGGTTCTGTGAGAGATTGATGGAATTAACCAAGATTTTGAAGACT 1080
 DB 1021 CAAGATCCCAAAAGGTTCTGTGAGAGATTGATGGAATTAACCAAGATTTTGAAGACT 1080
 QY 1081 CAATATCCGCTGTAACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1081 CAATATCCGCTGTAACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 1141 AGCTTTATCACTCATGATGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 DB 1141 AGCTTTATCACTCATGATGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 QY 1201 TATGTTGAGAGTTCCCATATTTTGGTATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1201 TATGTTGAGAGTTCCCATATTTTGGTATGATGATGATGATGATGATGATGATGATGATGAT 1260
 QY 1261 AGAGCAGCTGTAAATAAATTTCAAACTATGACAAAGCAAGATTTTCTGAGGCTTT 1320
 DB 1261 AGAGCAGCTGTAAATAAATTTCAAACTATGACAAAGCAAGATTTTCTGAGGCTTT 1320
 QY 1321 GAGAACAGTCAATTCAGATTCCTCTTAAAGAAAGTCTATGATGATGATGATGATGATGAT 1380
 DB 1321 GAGAACAGTCAATTCAGATTCCTCTTAAAGAAAGTCTATGATGATGATGATGATGATGAT 1380
 QY 1381 CCATGATCAACCTGTAAAGCCCTAGATGAGAGCTTCTGATGAGATTTGTGATGCG 1440
 DB 1381 CCATGATCAACCTGTAAAGCCCTAGATGAGAGCTTCTGATGAGATTTGTGATGCG 1440
 QY 1441 CCAGAAAGAGCCAGAGACCTGCGATCAGCTGCCCAATGACCTGACCTGGTTCAGACCTA 1500
 DB 1441 CCAGAAAGAGCCAGAGACCTGCGATCAGCTGCCCAATGACCTGACCTGGTTCAGACCTA 1500
 QY 1501 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 DB 1501 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 QY 1561 AAAATGTTTTTATTTTCTGTCGTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
 DB 1561 AAAATGTTTTTATTTTCTGTCGTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
 QY 1621 ATAGATCTTTTCAAAATTCAGAAAGACCTG 1650
 DB 1621 ATAGATCTTTTCAAAATTCAGAAAGACCTG 1650

RESULT 2
AA295199
ID AA295199 standard; DNA; 2092 BP.
XX
AC AA295199;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B4 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN W0200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; AAY78933.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 34-36; 72pp; English.
XX
XX This sequence represents the human UDP-glucuronosyltransferase 2B4
CC (UGT2B4) gene. UDP-glucuronosyltransferase (UGTs) are a family of
CC enzymes that catalyze the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may affect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see AA295051-Z95110). Probes which detect the
CC UGT2B locus polymorphisms can be used to detect altered UGT2B metabolism
CC of a substrate in an individual. The nucleic acid molecules comprising a
CC human UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of
CC UGT2B substrate, potential drug-drug interactions and adverse side
CC effects. The polymorphisms can be used as single nucleotide polymorphisms
CC (SNPs) for detecting genetic linkage related to phenotypic variation in
CC activity or expression of UGT2B protein. The polymorphism containing
CC nucleic acid molecules may also be used for generating genetically
CC modified non-human animals and for obtaining site specific gene
CC modification in cell lines.
XX
XX Sequence 2092 BP; 639 A; 398 C; 438 G; 617 T; 0 other;
SQ

Query Match 87.6%; Score 1445.4; DB 21; Length 2092;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1 AGCAACTGGAAAAACAGCATTCGATCGATCAGATGTCATGAAATGGAAGTCTTCAAGCTCT 60
Db 4 AGCAACTGGAAAAACAGCATTCGATCGATCAGATGTCATGAAATGGAAGTCTTCAAGCTCT 63
QY 61 TCTCTGATACAGCTGAGCTGTACTTACGCTCTGGAGTGTGGAAAGGCTGCGGTG 120
Db 64 TCTGCTGATACAGCTGAGCTGTACTTACGCTCTGGAGTGTGGAAAGGCTGCGGTG 123
QY 121 GCCCAGAAATTCAGCCACTGGATGATATTAAGCAATCCGTGATGAATTCGTCAGAG 180
Db

Db 124 GCCCAGAAATTCAGCCACTGGATGATATTAAGCAATCCGTGATGAATTCGTCAGAG 183
QY 181 AGGTCATGAGGTGACGTATTTGGCATCTTCAGCTTCATTTCTTTCGATCCCAACAGCCC 240
Db 184 AGGTCATGAGGTGACGTATTTGGCATCTTCAGCTTCATTTCTTTCGATCCCAACAGCCC 243
QY 241 ATCTACTCTTAAATTTGAACTTTATCTGTATCTTTAACTTAAATCTGAGTTTGAGATAT 300
Db 244 ATCTACTCTTAAATTTGAACTTTATCTGTATCTTTAACTTAAATCTGAGTTTGAGATAT 303
QY 301 TATCAACAGCTGTTTAAGAGATGGGCAAACTTCCAAAGACACATTTTGGTATATTT 360
Db 304 TATCAACAGCTGTTTAAGAGATGGGCAAACTTCCAAAGACACATTTTGGTATATTT 363
QY 361 TTCAACAGTCAAGAAATCATGTGGAATTTAATGATCATCTTAAGAAAGTTCTGTAAGA 420
Db 364 TTCAACAGTCAAGAAATCATGTGGAATTTAATGATCATCTTAAGAAAGTTCTGTAAGA 423
QY 421 TATAGTTTCAAAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 480
Db 424 TATAGTTTCAAAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 483
QY 481 TGCAGATGCTGTTTCCCTTTGGTGAGCTGCTGGCCGAGTTACTTAAATACCTTTGT 540
Db 484 TGCAGATGCTGTTTCCCTTTGGTGAGCTGCTGGCCGAGTTACTTAAATACCTTTGT 543
QY 541 CTACAGCTCCGCTTCTCTGCTGCTACGCAATTTGAAGCATATGAGAGAGCTTCTGTT 600
Db 544 CTACAGCTCCGCTTCTCTGCTGCTACGCAATTTGAAGCATATGAGAGAGCTTCTGTT 603
QY 601 CCTCTCTTCTTATGCTGCTGTTTATGATGCAAGTCAAGTACCAATGACCTTTCATAGA 660
Db 604 CCTCTCTTCTTATGCTGCTGTTTATGATGCAAGTCAAGTACCAATGACCTTTCATAGA 663
QY 661 GAGGTAATAAATATGATCTATGCTTATTTGAATTTTGGTCCAAATATTTGACAT 720
Db 664 GAGGTAATAAATATGATCTATGCTTATTTGAATTTTGGTCCAAATATTTGACAT 723
QY 721 GAAGAAATGGATACGTTTCAAGTGAAGTCTTGAAGAAAGCCCTACGTTATCTGAGAC 780
Db 724 GAAGAAATGGATACGTTTCAAGTGAAGTCTTGAAGAAAGCCCTACGTTATCTGAGAC 783
QY 781 AATGGCAAAAGCTGACATATGAGCTTATTCGAAACTACTGGAATTTCAATTTCTTCACC 840
Db 784 AATGGCAAAAGCTGACATATGAGCTTATTCGAAACTACTGGAATTTCAATTTCTTCACC 843
QY 841 ACTCTTACCAATGTTGAGTTCTGTTGAGAGACTTCACCTGCAAACTGCCAAACCTTACC 900
Db 844 ACTCTTACCAATGTTGAGTTCTGTTGAGAGACTTCACCTGCAAACTGCCAAACCTTACC 903
QY 901 GAAGAAATGGAAAGTGTTCAGAGCTCTGAGAAAGTGTGTGTTGTTTCTCT 960
Db 904 GAAGAAATGGAAAGTGTTCAGAGCTCTGAGAAAGTGTGTGTTGTTTCTCT 963
QY 961 GGGGTGATGGTCACTAAGCAAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC 1020
Db 964 GGGGTGATGGTCACTAAGCAAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC 1023
QY 1021 CAAGATCCCAAAAAGGTTCTGTGAGATTTGATGGAATTAACCAAGATCTTTAGAACT 1080
Db 1024 CAAGATCCCAAAAAGGTTCTGTGAGATTTGATGGAATTAACCAAGATCTTTAGAACT 1083
QY 1081 CAATATCTGGCTGTACAAAGTGAATCCCGAAGATGATCTTCTGGTATCCCAAAACCAA 1140
Db 1084 CAATATCTGGCTGTACAAAGTGAATCCCGAAGATGATCTTCTGGTATCCCAAAACCAA 1143
QY 1141 AGCTTTATCACTATGTTGGAATGATGGAATCTATGAGCTATTTACCATGGGCTCC 1200
Db 1144 AGCTTTATCACTATGTTGGAATGATGGAATCTATGAGCTATTTACCATGGAATCTCC 1203
QY 1201 TATGTTGGAGTTCCCATATTTTGGTGTACGCTTGATTAACATAGCTCACATGAAGGCCAA 1260
Db 1204 TATGTTGGAGTTCCCATATTTTGGTGTACGCTTGATTAACATAGCTCACATGAAGGCCAA 1263


```

Qy 361 TTCAAGATACAGAAATCATGTGAGCATTTAATGACATCTAGAAAAGTTCTGTAAAGA 420
Db 364 TTCAAGATACAGAAATCATGTGAGCATTTAATGACATCTAGAAAAGTTCTGTAAAGA 423
Qy 421 TATAGTTTCAATAAGAACTTATGAGAAAATCAAGAGTCAAGATTGATGTTGTTCT 480
Db 424 TATAGTTTCAATAAGAACTTATGAGAAAATCAAGAGTCAAGATTGATGTTGTTCT 483
Qy 481 TGCAGATGCTGTTTCCCTTGTGAGCTGTGCGAGTACTTAATAAATACCTTTGT 540
Db 484 TGCAGATGCTGTTTCCCTTGTGAGCTGTGCGAGTACTTAATAAATACCTTTGT 543
Qy 541 CTACAGCTCCGCTTCTCTCTGAGTACCAATTGAAAAGATAGTGAGGACCTTCTGT 600
Db 544 CTACAGCTCCGCTTCTCTCTGAGTACCAATTGAAAAGATAGTGAGGACCTTCTGT 603
Qy 601 CCTCTCTTCTATGCTGCTGTTTATGTAGAACTAAGTACCAATGACTTTTCAATGA 660
Db 604 CCTCTCTTCTATGCTGCTGTTTATGTAGAACTAAGTACCAATGACTTTTCAATGA 663
Qy 661 GAGGTTAAAAATATGATCTATGCTTATTTGAAATTTGTTCCAAATATTTGACAT 720
Db 664 GAGGTTAAAAATATGATCTATGCTTATTTGAAATTTGTTCCAAATATTTGACAT 723
Qy 721 GAAGAAGTGGATCAGTTCTACAGTGAAGTCTTGAAGAAGCCACTAGCTTATCTGAGC 780
Db 724 GAAGAAGTGGATCAGTTCTACAGTGAAGTCTTGAAGAAGCCACTAGCTTATCTGAGC 783
Qy 781 AATGGCAAAAGCTGACATATGCTTATTTGAAATCTAGTGGATTTTCAATTTCTCACCC 840
Db 784 AATGGCAAAAGCTGACATATGCTTATTTGAAATCTAGTGGATTTTCAATTTCTCACCC 843
Qy 841 ACTCTTACCAATGTTGAGTTCGTGGAGGACTCCAGCTGCAAAACCTGCAAAACCTTACC 900
Db 844 ACTCTTACCAATGTTGAGTTCGTGGAGGACTCCAGCTGCAAAACCTGCAAAACCTTACC 903
Qy 901 GAAGAAGTGAAGTGAAGTTCAGAGCTCTGAGAAATAGTGTGTTGTTCTC 959
Db 904 GAAGAAGTGAAGTGAAGTTCAGAGCTCTGAGAAATAGTGTGTTGTTCTC 963
Qy 960 TGGGTCGATGTCAGTAAACAGCTGAGAAAGGAGCCAAATGATTAATGATCAGCCCTTG 1019
Db 964 TGGGTCGATGTCAGTAAACAGCTGAGAAAGGAGCCAAATGATTAATGATCAGCCCTTG 1023
Qy 1020 CCAAGATCCCAAAAGGTTCTGTGAGATTTGATGGGAAATTAACCAATACTTTAGAGC 1079
Db 1024 CCAAGATCCCAAAAGGTTCTGTGAGATTTGATGGGAAATTAACCAATACTTTAGAGC 1083
Qy 1080 TCAATACTCGGCTGTACAAAGTGGATACCCAGAAATGATCTTGTGTCATCCCAAAACCA 1139
Db 1084 TCAATACTCGGCTGTACAAAGTGGATACCCAGAAATGATCTTGTGTCATCCCAAAACCA 1143
Qy 1140 AAGCTTTTATCACTATGCTGAGTGAATGAGATCTATGAAAGCTTATTAATCAATGAGGCTCC 1199
Db 1144 AAGCTTTTATCACTATGCTGAGTGAATGAGATCTATGAAAGCTTATTAATCAATGAGGCTCC 1203
Qy 1200 CTAATGTTGGAGATTTCCCAATTTTGTGATCAGCTTGAATCACTACTCAAGAGGCA 1259
Db 1204 CTAATGTTGGAGATTTCCCAATTTTGTGATCAGCTTGAATCACTACTCAAGAGGCA 1263
Qy 1260 AAGGAGAGCTGTAGAAATTAATCAATATGATGACCAAGCAAGATTTACTGAGGCTT 1319
Db 1264 AAGGAGAGCTGTAGAAATTAATCAATATGATGACCAAGCAAGATTTACTGAGGCTT 1323
Qy 1320 TGAGAGACAGTCAATACCGATTCCTCTTATTAAGAGAAATGATGATTTATCAAGAAATTC 1379
Db 1324 TGAGAGACAGTCAATACCGATTCCTCTTATTAAGAGAAATGATGATTTATCAAGAAATTC 1383
Qy 1380 ACCATGATTAACCTGTAAAGCCCCCTTATGATGAGCAAGTCTTCTGATTCGAGTTTGCATGC 1439
Db 1384 ACCATGATTAACCTGTAAAGCCCCCTTATGATGAGCAAGTCTTCTGATTCGAGTTTGCATGC 1443
Qy 1440 GCCAAGAAAGGAGCAAGCACTGCGATCAGCTGCCATGACCTCAGTCTGAGCACT 1499

```

```

Db 1444 GCCAAGAAAGGAGCAAGCACTGCGATTCAGGCTTCCAGCCAGCACTCAGCTTCCAGTACC 1503
Qy 1500 ACTCTATAGATGATGATGGGTTCTGCTGACCTGTGAGGCAAGCTGATATTTCTGTTCA 1559
Db 1504 ACTCTATAGATGATGATGGGTTCTGCTGACCTGTGAGGCAAGCTGATATTTCTGTTCA 1563
Qy 1560 CAAATGTTTTTATTTTCTGTCMAAAATTTAATAAAGTATGAAAGATGAAAAAGAGG 1619
Db 1564 CAAATGTTTTTATTTTCTGTCMAAAATTTAATAAAGTATGAAAGATGAAAAAGAG 1620
Qy 1620 AATGATCTTTC 1631
Db 1621 AATGATCTTTC 1632

RESULT 5
AAL41490
ID AAL41490 standard; DNA, 1639 BP.
XX
AC AAL41490;
XX
DT 21-NOV-2002 (first entry)
XX
DE Drug metabolizing enzyme encoding DNA - 7493833CB1.
XX
KW Anti-HIV; antiarteriosclerotic; dermatological; cyostatic; thyromimetic;
KW osteopathic; thrombolytic; ophthalmological; antileptic; hepatotropic;
KW antidiarrhoeic; antineoplastic; vitamin; immunogenic; autoimmune;
KW drug metabolizing enzyme; DME; inflammatory; cancer; cirrhosis; dwarfism;
KW contact dermatitis; cell-proliferative; cancer; cirrhosis; dwarfism;
KW developmental; hypothyroidism; endocrine; osteoporosis; thrombosis;
KW diabetes; glaucoma; keratitis; metabolic; hyperlipidemia; diarrhoea;
KW cystic fibrosis; gastrointestinal; gastroenteritis; liver; hepatitis;
KW Reye's syndrome; exogenous compound; gene therapy; enzyme; human; gene;
KW ds.
XX
XX Homo sapiens.
XX
OS
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1590
XX FT /*tag= a
XX FT /product= "Drug metabolizing enzyme protein"
XX
XX MO20026654-A2.
XX
XX 29-AUG-2002.
XX
XX PF 14-FEB-2002; 2002MO-US04918.
XX
XX PR 16-FEB-2001; 2001US-269643P.
XX PR 23-FEB-2001; 2001US-271333P.
XX PR 16-MAR-2001; 2001US-276767P.
XX PR 06-APR-2001; 2001US-282077P.
XX PR 19-APR-2001; 2001US-285447P.
XX PR 27-APR-2001; 2001US-287060P.
XX PR 03-MAY-2001; 2001US-288543P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Astronomoff A, Au-Yang J, Baughn MR, Ding L, Duggan BM;
XX Foreythe J, Gietzen KJ, Griffen JA, Lee EA, Lu Y, Richardson TW;
XX Ring HZ, Sanjanwala MM, Swarnakar A, Walla NK, Warren BA, Xu Y,
XX Yue H, Zebardjian Y;
XX
XX WPI; 2002-674949/72.
XX P-PsDB; AAO22794.
XX
XX New drug metabolizing enzymes (DME) useful for diagnosing, treating and
XX preventing diseases or conditions associated with aberrant DME
XX expression, e.g. cancer, AIDS, arteriosclerosis, diabetes, glaucoma,
XX hepatitis, osteoporosis -
XX

```


PS Claim 5; Page 161-162; 166pp; English

The invention relates to an isolated polypeptide comprising 12 fully defined sequences of 81-615 amino acids given in the specification; a naturally occurring amino acid sequence at least 90% identical to, having 81-559 amino acids, at least 91% identical to a sequence of 529 amino acids, or at least 97% identical to a sequence of 615 amino acids, all given in the specification; or a biologically active or immunogenic fragment of the polypeptide. The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of a drug metabolising enzyme (DME), such as autoimmune/inflammatory (e.g. AIDS, atherosclerosis, contact dermatitis), cell-proliferative (e.g. cancer, cirrhosis), developmental (e.g. dwarfism, hypothyroidism), endocrine (e.g. osteoporosis, thrombosis, diabetes), eye (e.g. glaucoma), keratitis, metabolic (e.g. hyperlipidaemia, cystic fibrosis), gastrointestinal (e.g. gastroenteritis, diarrhoea), or liver (e.g. hepatitis, Reye's syndrome) disorders. These are also useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of DME. The DME or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds that specifically bind to or modulate the activity of the polypeptide. The microarray is useful in monitoring or measuring protein-protein interactions, drug-target interactions, and gene expression therapy. The polynucleotides of the invention can be used in gene therapy. This polynucleotide sequence represents the DNA encoding a drug metabolising enzyme of the invention.

SQ Sequence 1639 BP; 470 A; 316 C; 372 G; 481 T; 0 other;

Query Match	80.7%;	Score 1331.4;	DB 24;	Length 1639;
Database:

Matches 1431; Conservative 0; Mismatches 166; Indels 0;

Matches 1431; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY	35	ATGTCTATTAATAATGACCTTCAGCTCTCTCCGGAACAGCTGAGCTGTATCTTAAGCTCT	94
Db	1	ATGTCTATTAATAATGACCTTCAGCTCTCTCTGCTGATACACGCTGAGCTGTACTTAACTCT	60
QY	95	GGGAGTTGTGAAAGGTGCTGTGTGGCCACAGAAATTCAGCCACTGGATGAATATTAAG	154
Db	61	GGGAGTTGTGAAAGGTGCTGTGTGGCCACAGAAATTCAGCCACTGGATGAATATTAAG	120
QY	155	ACAATCTGTGAAGAACTGTCCAGAGAGGTCA TGAAGTACGTATATGGA TCTTCAGCT	214
Db	121	ACAATCTGTGAAGAACTGTCCAGAGAGGTCA TGAAGTACGTATATGGA TCTTCAGCT	180
QY	215	TCCATTTCTTGTGATCCCAAGACCCTACTCTTAATTTGAAGTTATCCGTATCT	274
Db	181	TCCATTTCTTGTGATCCCAAGACCCTACTCTTAATTTGAAGTTATCCGTATCT	240
QY	275	TTTACTTAAACCTGAGTTTGAAGATATTTATCAAGCACTGTGTTAAGATGGCGAACTT	334
Db	241	TTTACTTAAACCTGAGTTTGAAGATATTTATCAAGCACTGTGTTAAGATGGCGAACTT	300
QY	335	CCAAAAGACACATTTTGGTCACTATTTTTCACAAGTACAAAGAAATCATGTGGACATTTAAT	394
Db	301	CCAAAAGACACATTTTGGTCACTATTTTTCACAAGTACAAAGAAATCATGTGGACATTTAAT	360
QY	395	GACATATCTTAGAAAGTCTGTAAAGATATAGTTTCAATTAAGAAACTTATGAAAGAACTA	454
Db	361	GACATATCTTAGAAAGTCTGTAAAGATATAGTTTCAATTAAGAAACTTATGAAAGAACTA	420
QY	455	CAGGATCAAGATTGATGATGTTGTTCTTCAGATGCTGTTTCCCTTTGGTGGAGCTGCTG	514
Db	421	CAGGATCAAGATTGATGATGTTGTTCTTCAGATGCTGTTTCCCTTTGGTGGAGCTGCTG	480
QY	515	GCCGAGTTACTTAAATATACCCCTTTGTCTACAGCCTCCGCTCTCTCCGTGCTACGCAATT	574
Db	481	GCTGAGCTATTTAACAATACCTTTGTGTAAGTGCACAGCTCTCTCCGTGCTACACATTTT	540
QY	575	GAAGAAGATAGTGAAGGACTTCTGTTCCCTCTTCTCTATGTGCTGTTGTTATGTACAAA	634

AD45991 standard; cDNA; 1991 BP.
AD45991;
27-DEC-2002 (first entry)
Human UGT2B7 cDNA.
Human, UDP-glucuronosyl transferase; UGT; UGT2B7; toxicity; cancer;
therapy; epirubicin; cytostatic; enzyme; gene; ss.
Homo sapiens.
Key Location/Qualifiers
CDS 151..1740
/*tag= a
/product= "Human UGT2B7 protein"
MO200259375-A2.
01-AUG-2002.
25-JAN-2002; 2002MO-US02083.
26-JAN-2001; 2001US-264534P.
(UTCH-) UNIV CHICAGO.
Ratain MJ, Innocenti F, Das S, Iyer L, Sawyer M;
WPI; 2002-691534/74.
P-PSDB; AAE26617.
Determining the dose of a UGT2B7-glucuronidated drug for treating
cancer, comprises determining the level of UGT2B7 activity or
expression in a patient
Disclosure: Page 143-144; 160pp; English.
The invention relates to an UDP-glucuronosyl transferase (UGT) enzyme,
UGT2B7. The invention also relates to compositions and methods for
optimizing UGT2B7 substrate dosings and for predicting UGT2B7 substrate
toxicity. The method is useful in determining the dose of a UGT2B7-
glucuronidated drug that may be used in treating cancer patients. It
is also useful in determining persons at risk for epirubicin toxicity,
in reducing or eliminating side effects associated with epirubicin
treatment, and in ways of increasing the efficacy of dosage regimens.
The present sequence is human UGT2B7 cDNA.
SQ Sequence 1991 BP; 618 A; 357 C; 419 G; 597 T; 0 other;
Query Match 77.1%; Score 1272.6; DB 24; Length 1991;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 1407; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
QY 1 AGCACTGGAAGAAAGCAAGCTTGCATTCAGATGCTGTATGAAATGAGCTTCACTCT 60
DB 117 AGCACTGGAAGAAAGCAAGCTTGCATTCAGATGCTGTATGAAATGAGCTTCACTCT 176
QY 61 TTCCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGAAAGTCTGGTGTG 120
DB 177 TTTCGTAATACAGCTGAGCTTGTTCGTTAGCTCTGGAGTTGTGAAAGTCTGGTGTG 236
QY 121 GCCCAGAGATTCAGGCACTGATGATATTAAGACATCTGATGAACTTGTCCAGAG 180
DB 227 GGCAGAGATTCAGGCACTGATGATATTAAGACATCTGATGAACTTGTCCAGAG 296
QY 181 AGGTGATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 297 AGGTGATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 356
QY 241 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

DB 357 ATCCGCTCTTAAATGAAATTTATCCACATCTTTAATCTAATAAATGAGTGGAAATTT 416
QY 301 TATCAGAGCTGGTGTAGAGATGGCAGAACTTCAAAAGACATTTTGTGATATTT 360
DB 417 CATCATGCAACAGATTAGAGATGCTCAGACCTTCCAAAAGATACATTTTGTGATATTT 476
QY 361 TTCAGAGTACAGAAATTCATGTCAGATTTAATGACATCTTGAAGATTTCTGTAAGA 420
DB 477 TTCAGAGTACAGAAATTCATGTCAGATTTAATGACATCTTGAAGATTTCTGTAAGA 536
QY 421 TATGATTTCAATTAAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCT 480
DB 537 TGTAGTTTCAATTAAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCT 596
QY 481 TGCAGATGCTGTTTCCCTTGTGAGCTGCTGGCCAGTTACTTAAATACCTTTGT 540
DB 597 TGCAGATGCTGTTTCCCTTGTGAGCTGCTGGCCAGTTACTTAAATACCTTTGT 656
QY 541 CTACAGCTCCGCTTCTCTCTGCTACGCAATTTGAAAGCATAGTGAGACTTCTGT 600
DB 657 GTACAGCTCAGCTTCTCTCTGCTACGCAATTTGAAAGCATAGTGAGACTTCTGT 716
QY 601 CCTCTCTTCTATGCTGCTGTTGATGTCAGAACTAAGTACCAATGACTTTCATGA 660
DB 717 CCTCTCTTCTATGCTGCTGTTGATGTCAGAACTAAGTACCAATGACTTTCATGA 776
QY 661 GAGGTTAAATTAATGATCTATGCTTATTTGAAATTTGTTCCAAATATTTGACAT 720
DB 777 GAGGTTAAATTAATGATCTATGCTTATTTGAAATTTGTTCCAAATATTTGACAT 836
QY 721 GAAAGATGGGATCAGTTCTACAGTGAAGTTCTAGAAAGACCACTACGTTATCTGAG 780
DB 837 GAAAGATGGGATCAGTTCTACAGTGAAGTTCTAGAAAGACCACTACGTTATCTGAG 896
QY 781 AATGGCAAAAGCTGACATATGCTTATTTCAAACTACTGAGATTTTCAATTTCTCAC 840
DB 897 AATGGGAAAGCTGACATATGCTTATTTCAAACTACTGAGATTTTCAATTTCTCAC 956
QY 841 ACTCTTACCAATGATGAGTGTGTTGAGAGCTCACTGCAAACTGCAAACTGCAAACT 900
DB 957 ACTCTTACCAATGATGAGTGTGTTGAGAGCTCACTGCAAACTGCAAACTGCAAACT 1016
QY 901 GAAAGAAATGGAAGATTTTTCAGAGCTCTGGAAGAAATGATGATGATGATGATGAT 960
DB 1017 TAAAGAAATGGAAGATTTTTCAGAGCTCTGGAAGAAATGATGATGATGATGATGAT 1076
QY 961 GGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 1077 GGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1136
QY 1021 CAAGATCCCAAAAAGTTCTGTGAGATTTGATGGAATTAACAGATCTTGAAGACT 1080
DB 1137 CAAGATCCCAAAAAGTTCTGTGAGATTTGATGGAATTAACAGATCTTGAAGACT 1196
QY 1081 CAATCTGCTGATCAAGTGAATCCCAAGATGATCTTGTGATCTCCCAAAACCA 1140
DB 1197 CAATCTGCTGATCAAGTGAATCCCAAGATGATCTTGTGATCTCCCAAAACCA 1256
QY 1141 AGCTTTATCACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1257 AGCTTTATCACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1316
QY 1201 TATGTTGGAGATTTCCATTTTGTGATCAGCTGATTAACATGATCAGTGAAGGCCAA 1260
DB 1317 TATGTTGGAGATTTCCATTTTGTGATCAGCTGATTAACATGATCAGTGAAGGCCAA 1376
QY 1261 AGAGCAGCTGTAGAAATTAATCTTCAAACTATGCAAGCAAGATTTTCTAGAGGCTTT 1320
DB 1377 GGGAGCAGCTGTAGAAATTAATCTTCAAACTATGCAAGCAAGATTTTCTAGAGGCTTT 1436
QY 1321 GAGAACTGATCAATTCAGATCTCTTATTAAGAGATCTATGATTAATCAAGATTTCA 1380
DB 1437 GAGAACTGATCAATTCAGATCTCTTATTAAGAGATCTATGATTAATCAAGATTTCA 1496


```

Db      ||| 961 TGACAGAAAGAGGCCCACTAATGTGATGAGCCCTGGCCACATCCCAAAAGGTTCC 1020
Qy      ||| 1041 TGTGAGATTGATGGAGATTAACCCAGATCTTTAGACTCAATATCTGGCTGTACAACT 1100
Db      ||| 1021 TGTGAGATTGATGGAGATTAACCCAGATCTTTAGACTCAATATCTGGCTGTACAACT 1080
Qy      ||| 1101 GGATGATCCCAAGATCTTGTGTATCCCAAAACCAAGCTTTATCACTAGTGTG 1160
Db      ||| 1081 GGATGATCCCAAGATCTTGTGTATCCCAAAACCAAGCTTTATCACTAGTGTG 1140
Qy      ||| 1161 GAATGATGGAGATCTATGAGATCTTTACATGGGGTCCCTATGGTGGAGTCCCATAT 1220
Db      ||| 1141 GAGGCAATGGATCTACGAGGCAATCTACATGGGATCCATATGGTGGATTCATGT 1200
Qy      ||| 1221 TTGGTATGATCTGATTAACATAGCTCATAGAAAGCCAAAGAGACAGCTGTAGAAATTA 1280
Db      ||| 1201 TTGGCGATCAACCTGATTAACATAGCTCATAGAAAGCCAGGAGACAGCTGTAGAGTGG 1260
Qy      ||| 1281 ACTTCAAAACCTATGACAAAGAGATTTACTGAGGGCTTTGAGAAACAGTCAATACCGAT 1340
Db      ||| 1261 ACTTCAACACATGTCAGATACAGCTTGTGATGCTTGAAGAGATTAATTAATGATC 1320
Qy      ||| 1341 CCTCTTAAAGAGATCTATGAGATTAACAGATTCACCATGATCAACCTGTAAAGC 1400
Db      ||| 1321 CTTCAATAAAGAGATTTATGAATTAATCAAGATTCACATGATCAACGATGAAGC 1380
Qy      ||| 1401 CCTGATGAGACAGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1460
Db      ||| 1381 CCTGATGAGACAGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Qy      ||| 1461 TGGCATGAGCTCCATGACCTCAGCTGCTCCAGCACTACTCTATGATGATGATGATGATGATGAT 1520
Db      ||| 1441 TTGGGTTGACGCCACACACCTCAGCTGCTCCAGCACTACTCTATGATGATGATGATGATGAT 1500
Qy      ||| 1521 TCTGCTGACCTGTTGTCACCTGCTATATCTTGTTCACAAATGTTTTTATTTTCT 1580
Db      ||| 1501 TCTGCTGCTGCTGTTGTCACCTGCTATATCTTGTTCACAAATGTTTTTATTTTCT 1560
Qy      ||| 1581 GTCAAAATTTATTAATAAATAAGAAAGATGAAGAGAGGAAATGATCTTTC 1611
Db      ||| 1561 TCTGGAAGTTGCTGAAAGAAAGAGGAAAGAAATGATGATGATGATGATGATGATGAT 1611

RESULT 8
ABL68501
ID      ABL68501 standard; DNA; 1855 BP.
XX      AC
XX      ABL68501;
XX      15-MAY-2002 (first entry)
DE      Kidney cancer related gene sequence SEQ ID NO:6838.
XX      Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM      stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM      cytotoxic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
KM      gene; ds.
XX      Homo sapiens.
XX      OS
XX      WO200194629-A2.
XX      PD
XX      13-DEC-2001.
XX      PF
XX      30-MAY-2001; 2001WO-US10838.
XX      PR
XX      05-JUN-2000; 2000US-209473P.
XX      05-JUN-2000; 2000US-209531P.
XX      18-SEP-2000; 2000US-23133P.
XX      18-SEP-2000; 2000US-231617P.
XX      20-SEP-2000; 2000US-234009P.
XX      20-SEP-2000; 2000US-234034P.

```

```

PR      20-SEP-2000; 2000US-234052P.
PR      22-SEP-2000; 2000US-234509P.
PR      22-SEP-2000; 2000US-234567P.
PR      25-SEP-2000; 2000US-234923P.
PR      25-SEP-2000; 2000US-234924P.
PR      25-SEP-2000; 2000US-235077P.
PR      25-SEP-2000; 2000US-235082P.
PR      25-SEP-2000; 2000US-235134P.
PR      25-SEP-2000; 2000US-235280P.
PR      26-SEP-2000; 2000US-235637P.
PR      26-SEP-2000; 2000US-235638P.
PR      26-SEP-2000; 2000US-235639P.
PR      27-SEP-2000; 2000US-235711P.
PR      27-SEP-2000; 2000US-235720P.
PR      27-SEP-2000; 2000US-235840P.
PR      27-SEP-2000; 2000US-235863P.
PR      28-SEP-2000; 2000US-236028P.
PR      28-SEP-2000; 2000US-236032P.
PR      28-SEP-2000; 2000US-236033P.
PR      28-SEP-2000; 2000US-236034P.
PR      28-SEP-2000; 2000US-236109P.
PR      28-SEP-2000; 2000US-236111P.
PR      28-SEP-2000; 2000US-236842P.
PR      29-SEP-2000; 2000US-236891P.
PR      29-SEP-2000; 2000US-237172P.
PR      02-OCT-2000; 2000US-237173P.
PR      02-OCT-2000; 2000US-237278P.
PR      02-OCT-2000; 2000US-237294P.
PR      02-OCT-2000; 2000US-237295P.
PR      02-OCT-2000; 2000US-237316P.
PR      03-OCT-2000; 2000US-237425P.
PR      03-OCT-2000; 2000US-237598P.
PR      03-OCT-2000; 2000US-237604P.
PR      03-OCT-2000; 2000US-237606P.
PR      03-OCT-2000; 2000US-237608P.
PR      01-NOV-2000; 2000US-244867P.
PR      01-NOV-2000; 2000US-245084P.

(AVAL- ) AVALON PHARM.
XX      PA
XX      Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX      Soppet DR, Weaver Z;
XX      WPI; 2002-188264/24.
XX      DR
XX      Screening for anti-neoplastic agent involves exposing cells to a
XX      chemical agent to be tested for anti-neoplastic activity, and
XX      determining a change in expression of a gene of a signature gene set
XX      Claim 1; SEQ ID 6838; 44pp; English.
XX      PS
XX      The present invention describes a method (M1) for screening for an
XX      anti-neoplastic agent. The method involves exposing cells to a chemical
XX      agent to be tested for anti-neoplastic activity, determining a change in
XX      expression of at least one gene (I) of a signature gene set, where (I)
XX      comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX      to ABL70110), or is at least 95% identical to (S), where a change in
XX      expression is indicative of anti-neoplastic activity. (I) has cytotoxic
XX      activity and can be used in gene therapy. M1 can be used for screening
XX      an anti-neoplastic agent, and can be used for producing a product which
XX      is the data collected with respect to the anti-neoplastic agent as a
XX      result of M1, and the data is sufficient to convey the chemical
XX      structure and/or properties of the agent. M1 can be used in the
XX      treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX      oesophageal, ovarian, kidney, prostate or pancreatic cancer.
XX      CC      adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX      CC      infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX      CC      carcinoma, papillary carcinoma and Wilms tumour.
XX      SQ
XX      Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;

Query Match      75.9%; Score 1252.6; DB 24; Length 1855;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```

QY 21 TGCATGATCAGGATGCTATGAAATGAGCTTCTCTGATACAGCTGAGCT 80
 Db 1 TGCATGACACGAGATGCTGTGAAATGAGCTTCTGATATTTTGTATTAACATGAGCT 60
 QY 81 GTTACTTAACTCTGGGAGTTGTGGAAAGTGTGTGTGGCCACAGAAATTCAGCACT 140
 Db 61 TTTGCTTAACTCTGGGAGTTGTGGAAAGTGTGTGTGGCCACAGAAATTCAGCACT 120
 QY 141 GGATGATATTAACAACTCTGATGAACTTGTCCAGAGAGTGTATGAGTGTAT 200
 Db 121 GGAATGATATTAACAACTCTGATGAACTTGTCCAGAGAGTGTATGAGTGTAT 180
 QY 201 TGGGATCTTCACTTCTTCTGATCCCAAGCCATCTACTTAAATTTTGAAG 260
 Db 181 TGGGATCTTCACTTCTTCTGATCCCAAGCCATCTACTTAAATTTTGAAG 240
 QY 261 TTTATCTGTATCTTTAACTAAACCTGATTTGAGATATTATCAAGCAGCTGTAGA 320
 Db 241 TTTATCTGTATCTTTAACTAAACCTGATTTGAGATATTATCAAGCAGCTGTAGA 300
 QY 321 GATGGGAGAACTTCCAAAAGACATTTTGTGATATTTTTCAGAGTACAAAGATCA 380
 Db 301 GATGGGAGAACTTCCAAAAGACATTTTGTGATATTTTTCAGAGTACAAAGATCA 360
 QY 381 TGTGACATTTAATGACATCTTAAAGATCTGTAAAGATATAGTTTCAATAGAAAC 440
 Db 361 TGTGACATTTAATGACATCTTAAAGATCTGTAAAGATATAGTTTCAATAGAAAC 420
 QY 441 TTATGAGAAACCTACAGAGTCAAGATTTGATTTGTCTTGCAGATGCTGTTCCT 500
 Db 421 TTATGAGAAACCTACAGAGTCAAGATTTGATTTGTCTTGCAGATGCTGTTCCT 480
 QY 501 TTGTGAGCTGTGTGGCCGAGTACTTAAATACCTTTGTCTTAAGCTCCGCTTCTCTC 560
 Db 481 GTATGAGCTGTGTGGCCGAGTACTTAAATACCTTTGTCTTAAGCTCCGCTTCTCTC 540
 QY 561 CTGGCTACGCAATTTGAAAGCATAGTGGAGACTTCTGTCCTCTTCTGATAGTGCCTG 620
 Db 541 CTGGCTACGCAATTTGAAAGCATAGTGGAGACTTCTGTCCTCTTCTGATAGTGCCTG 600
 QY 621 TTGTATGTCAGAACTAAGTACCAATGACTTTTCAATAGAGGGTAAATAATATGATCT 680
 Db 601 TTGTATGTCAGAACTAAGTACCAATGACTTTTCAATAGAGGGTAAATAATATGATCT 660
 QY 681 ATGTGCTTATTTGAAATTTGTGTCAAATATTTGATGATGAAAGTGGATGATTTCT 740
 Db 661 ATGTGCTTATTTGAAATTTGTGTCAAATATTTGATGATGAAAGTGGATGATTTCT 720
 QY 741 ACAGTGAAGTCTAGGAAAGCCACTAGTATCTGAGACAAATGGCAAAAGCTGACATAT 800
 Db 721 ATAGTGAAGTCTAGGAAAGCCACTAGTATCTGAGACAAATGGCAAAAGCTGACATAT 780
 QY 801 GGCTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCTTACCAATTTTGAAT 860
 Db 781 GGCTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCTTACCAATTTTGAAT 840
 QY 861 TCGTTGAGAGCTCCTCACTGCAAACTGTGCAAACTTACCGAGAGAAATGGAGAGTTTG 920
 Db 841 TCGTTGAGAGCTCCTCACTGCAAACTGTGCAAACTTACCGAGAGAAATGGAGAGTTTG 900
 QY 921 TCCAGAGCTCTGAGAAATGGTGTGTGTGTTTCTCTGGGGTCAATGCTGATACA 980
 Db 901 TACAGAGCTCTGAGAAATGGTGTGTGTGTTTCTCTGGGGTCAATGCTGATACA 960
 QY 981 CGTCAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGATCCCAAGAGGTTTC 1040
 Db 961 TGACAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGATCCCAAGAGGTTTC 1020
 QY 1041 TGTGAGATTTGATGGGAAATTAACCAAGATCTTTAGACTCAATACTGCTGTACAAGT 1100
 Db 1021 TGTGAGATTTGATGGGAAATTAACCAAGATCTTTAGACTCAATACTGCTGTACAAGT 1080

QY 1101 GGATACCCAGAAATGATCTTCTGATCCCAAAACCAAGCTTTTATCAGTATG 1160
 Db 1081 GGATACCCAGAAATGATCTTCTGATCCCAAAACCAAGCTTTTATCAGTATG 1140
 QY 1161 GAATGAAATGGGATCTATGAAGCTATTTTACATGGGGTCCCTATGTGGAGTCCATAT 1220
 Db 1141 GAGCAATGGGATCTATGAAGCTATTTTACATGGGGTCCCTATGTGGAGTCCATAT 1200
 QY 1221 TTGTGATCAGCTTGAATTAATAGCTCAATGAAAGGCAAGAGAGCTGTAGAATA 1280
 Db 1201 TTGTGATCAGCTTGAATTAATAGCTCAATGAAAGGCAAGAGAGCTGTAGAATA 1260
 QY 1281 ACTTCAAACTATGACAAAGGAGATTTTCTGAGGGCTTTGAGAACAGTACCAT 1340
 Db 1261 ACTTCAAACTATGACAAAGGAGATTTTCTGAGGGCTTTGAGAACAGTACCAT 1320
 QY 1341 CTTCTTAAAGAGATGCTATGAGATTTATCAAGAAATTCACATGATCAACCTGTAAAG 1400
 Db 1321 CTTCTTAAAGAGATGCTATGAGATTTATCAAGAAATTCACATGATCAACCTGTAAAG 1380
 QY 1401 CCTAGATGAGAGCTTCTGATGAGTGTGTGATGAGCCCAAGAGCCAGAGACC 1460
 Db 1381 CCTAGATGAGAGCTTCTGATGAGTGTGTGATGAGCCCAAGAGCCAGAGACC 1440
 QY 1461 TGGGATCAGCTGCCATGACCTCACCTGTTCCAGACACTCTATAGATGTGATGGT 1520
 Db 1441 TGGGATCAGCTGCCATGACCTCACCTGTTCCAGACACTCTATAGATGTGATGGT 1500
 QY 1521 TCTGTGACCTGTGTGGCACTGTATATCTTGTGACAAATGTTTAAATTTTCT 1580
 Db 1501 TCTGTGACCTGTGTGTGGCACTGTATATCTTGTGACAAATGTTTAAATTTTCT 1560
 QY 1581 GTCAAAATTTTAAATCTAAGAAATGAGAAAGAGGAAATGATTTTCT 1631
 Db 1561 TGTGAAATTTTGTGTAGAAAGCAAGAAAGGAAATGATTTTATATC 1611

RESULT 9
 ABL68868
 ID ABL68868 standard, DNA, 1855 BP.
 XX
 AC ABL68868;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Kidney cancer related gene sequence SEQ ID NO:7205.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytosolic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001MO-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.


```

Db 781 GGCCTTATCGAAACTCTGGAATTTTCAGTTTCATATCCACTCTTACCAATGTTGAT 840
Qy 861 TCGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGTTTG 920
Db 841 TTGTTGGAGACTCCACTGCAAACTGCAAAACCCCTCTTACGGAATGGAAGCTTTG 900
Qy 921 TCCAGAGCTCTGAGAAAATGTTGTTGTTTCTCTGGGTCGATGTCAGTAACA 980
Db 901 TACAGAGCTCTGAGAAAATGTTGTTGTTTCTCTGGGTCGATGTCAGTAACA 960
Qy 981 CGTCAGAAAAGAGGCCCAATGTAATTCATGATCAGCCCTTGCCAAAGTCCCAAAAAGTTTC 1040
Db 961 TGACAGAAAAGAGGCCCAATGTAATTCATGATCAGCCCTTGCCAAAGTCCCAAAAAGTTTC 1020
Qy 1041 TGTGAGATTTGATGGAATTAACAGATTAATTCAGTCAATCTGCGGCTGTAAGT 1100
Db 1021 TGTGAGATTTGATGGAATTAACAGATTAATTCAGTCAATCTGCGGCTGTAAGT 1080
Qy 1101 GGATACCCGAGATATCTTCTGTTGTCATCCCAAAACCAAAAGCTTTTATCACTCATGTTG 1160
Db 1081 GGATACCCGAGATATCTTCTGTTGTCATCCCAAAACCAAAAGCTTTTATCACTCATGTTG 1140
Qy 1161 GAATGAATGGATCTATGAAGCTATTACATGAGGCTCCATGTTGGAGTTCCTCATAT 1220
Db 1141 GAGCCAAATGGATCTATGAAGCTATTACATGAGGCTCCATGTTGGAGTTCCTCATAT 1200
Qy 1221 TTGGATCAGCTTATTAACATGATCTCATGAGGCAAGGAGGAGCAGCTGTGAGAAATA 1280
Db 1201 TTGGATCAGCTTATTAACATGATCTCATGAGGCAAGGAGGAGCAGCTGTGAGAAATA 1260
Qy 1281 ACTTCAAAATCTATGACAAAGCAAGATTTACTGAGGCTTTTGAACAGTCAATACCAT 1340
Db 1261 ACTTCAAAATCTATGAGTACAGTACAGCTGCTGAATGCAATGAAGAGTAATTAATGATC 1320
Qy 1341 CCTCTTAATGAAGATGCTATGATTAATCAAGATTCACATGATCACTGTAAAGC 1400
Db 1321 CTTCAATTAATGAAGATGCTATGATTAATCAAGATTCACATGATCACTGTAAAGC 1380
Qy 1401 CCTGATGATGAGATCTTCTGATGAGTTCATGATGAGCAAGGAGGCAAGCACC 1460
Db 1381 CCTGATGATGAGATCTTCTGATGAGTTCATGATGAGCAAGGAGGCTAAACACC 1440
Qy 1461 TCGCATGAGCTCCCATGAGCTCACTGAGTTCAGACACTACTCTATAGATGATGGGT 1520
Db 1441 TTCCGGTTGCAACCCACAGCCTCACTGTTCCAGTACCACTTTGATGATGGGT 1500
Qy 1521 TCTGCTGAGCTGTGTGAGCACTGCTATATCTTGTTCACAAAATGTTTTATTTTCT 1580
Db 1501 TCTGCTGAGCTGTGTGAGCACTGCTATATCTTGTTCACAAAATGTTTTATTTTCT 1560
Qy 1581 GTCAAAAATTTTATTAATAGATTAAGATGAAGAGGAGATAGATCTTTC 1631
Db 1561 TCTGGAAGTTTCTGAGAAAAGCAAGAGGAGAAAATGATTAATTAATC 1611

```

RESULT 11
AAD06860
ID AAD06860 standard; DNA; 1722 BP.

AC AAD06860;
XX
XX
DT 06-AUG-2001 (first entry)
XX
DE Human breast cancer specific gene-2 (BCSG-2) #2.
XX
XX Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
XX breast cancer; therapeutic; gene therapy; ds.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 10..1599
XX FT
XX FT /*tag= a

```

FT /product= "Human BCSG-2 protein"
XX
XX MO200137779-A2.
XX
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-US32056.
XX
XX 23-NOV-1999; 99US-0166973.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Salceda S, Caffery R, Recipon H, Sun Y;
XX
XX WPI: 2001-367602/38.
XX
XX P-PSDB; AAE02188.
XX
XX Novel breast cancer specific gene for diagnosing, monitoring, staging,
XX imaging, preventing and treating cancers, particularly breast cancer
XX
XX Claim 1; Page 54-55; 66pp; English.
XX
XX The invention relates human breast cancer specific genes (BCSG's) and
XX their corresponding proteins. BCSG is useful for diagnosing, staging,
XX monitoring, imaging, preventing and treating breast cancers. BCSG is also
XX useful for inducing an immune response against a target cell expressing
XX BCSG. The invention also provide methods for detecting genetic lesions or
XX mutations in BCSG, thereby determining if a human with the genetic lesion
XX is at risk for breast cancer or has breast cancer. BCSG antibodies
XX labelled with paramagnetic ions or radioisotopes is useful for imaging
XX breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is
XX useful for treating breast cancer. BCSG is useful in the rational design
XX of new therapeutics for imaging and treating cancers. BCSG is also used
XX in gene therapy. The present DNA sequence is human breast cancer specific
XX gene-2 (BCSG-2) or Gene ID 480489.
XX
XX Sequence 1722 BP; 523 A; 319 C; 371 G; 509 T; 0 other;
XX
XX Query Match 75.0%; Score 1238; DB 22; Length 1722;
XX Best Local Similarity 85.7%; Pred. No. 0;
XX Matches 1376; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

```

```

Qy 506 GAGCTGCGCCGAGTACTTAATAAAATACCTTTGTCTACAGCTCCGCTCTCTCTGGC 565
Db 481 GAGCTGCGCCGAGTACTTAATAACATACGTTTGTGTACAGCTCCGCTTACTCTGGC 540
Qy 566 TACCGAATTGAAAAGCAATAGTGGAGACTTCTGTTCCTCCTTATGTCCTGTGTT 625
Db 541 TACCAATTTGAAAAGCAGAGTGGAGACTGATTTTCCCTCCTTCACTATCTATGTT 600
Qy 626 ATGTCAGAACTAAGTGAAGCAATGACTTTCATAGAGAGGTAAATAATATCATATG 685
Db 601 ATGTCAAAATTAAGTATCAATGACTTTCATAGAGAGGTAAATAATATCATATG 660
Qy 686 CTTTATTTTGAATTTGGTCCAAATATTTGACATAGAAGAGTGGATCACTTCACT 745
Db 661 CTTTATTTTGAATTTGGTCCAAATGCTGTATAGAAAGAGTGGATCACTTCACT 720
Qy 746 GAAGTTCTAGAGAAGCCCACTACGTTATCTGAGCAATGSCAAAAGCTGACATATG 805
Db 721 GAAGTTTAGAGAAGCCCACTACCTTATTTGAGCAATGGAAGAAAGTGAATATGCT 780
Qy 806 ATTGGAATCTATGGGATTTTCAATTTCTCAACCACTTCACTTCAAAATGTTGCT 865
Db 781 ATGCAAACTCCTGAGTTTCAATTTCTCACTTCACTTCAAAAGTTGATTTGTT 840
Qy 866 GGAGGACTCCCACTGCAAACTGSCAAAACCCCTACCGAAGAAATGAGAGTTGTCAG 925
Db 841 GGAGGATTCCTCACTGCAAACTGSCAAAACCCCTACCTAAGAAATGAGAGTTGTCAG 900
Qy 926 AGCTCTGAGAAAATGTTGTGTGTGTTTCTGTGGGTCGATGTCAGTAAACGTC 985
Db 901 AGCTCTGAGAAAATGTTGTGTGTGTTTCTGTGGGTCGATGTCAGTAAACGTC 960
Qy 986 GAAGAAAGGCAATGTAATGTCATCAGCCCTTGGCAAGATCCCAAAAAGTTCTG 1045
Db 961 GCAGAAAGGCAATGTAATGTCACACCCCTTGGCAAGATCCCAAAAAGTTCTG 1020
Qy 1046 AGATTGATGGGAATTAACCAAGATCTTAGAGACTCACTCGGCTGTACAGTATG 1105
Db 1021 AGATTGACGGGAATTAACCAAGATCTTAGAGACTCACTCGGCTGTACAGTATG 1080
Qy 1106 CCCAGAAATGATCTTGTGTATCCCAAAACCAAGCTTTTATCACTCATGTGATG 1165
Db 1081 CCCAGAAATGATCTTGTGTATCCCAAAACCAAGCTTTTATCACTCATGTGATG 1140
Qy 1166 AATGGATCTATGAAGCTATTTACATGGGTCCTTATGTCGAGTTCATATTTG 1225
Db 1141 AATGGATCTATGAAGCTATTTACATGGGTCCTTATGTCGAGTTCATATTTG 1200
Qy 1226 GATCAGCTGTATCATATGCTCAGATGAAAGCCAAAGAGAGAGCTGTAGAAATTA 1285
Db 1201 GATCAGCTGTATCATATGCTCAGATGAAAGCCAAAGAGAGAGCTGTAGAAATTA 1260
Qy 1286 AAAAATATGACAAAGGAAATTTACTGAGGCTTTGAGAACATGATTCAGATTCCT 1345
Db 1261 AAAAATATGACAAAGGAAATTTACTGAGGCTTTGAGAACATGATTCAGATTCCT 1320
Qy 1346 TATTAAGAAATGCTATGATATTAACAAGATTCACATGATCAACCTGTAAAGCC 1405
Db 1321 TATTAAGAAATGCTATGATATTAACAAGATTCACATGATCAACCTGTAAAGCC 1380
Qy 1406 GATGAGAGAGCTTCTGATGAGATTTGTCATGCGCCCAAGAGAGCCAAAGAGCT 1465
Db 1381 GATGAGAGAGCTTCTGATGAGATTTGTCATGCGCCCAAGAGAGCCAAAGAGCT 1440
Qy 1466 TCAGTGCCTCATGACCTGCTGTTCCAGACTCTATAGATGATGATGCTGCTG 1525
Db 1441 GTTGAGGCTCATGACCTGCTGTTCCAGACTCTATAGATGATGATGCTGCTG 1500
Qy 1526 CTGACCTGTGTGGCACTGCTATATTTCTGTTCAAAAATGTTTTATTTTCTGTC 1585
Db 1501 CTGACCTGTGTGGCACTGCTATATTTCTGTTCAAAAATGTTTTATTTTCTGTC 1560

```

```

Qy 1586 AATTTAATAAACTAGAAAGATGAAAGAGGAGGATATGATCTTTC 1631
Db 1561 AAGTTTCTAGAAAAGGAGAAAGAGGAAAAAGATTAATGTTATCTC 1606

RESULT 12
ABN97261
ID ABN97261 standard; DNA; 2799 BP.
XX
XX AC ABN97261;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Gene #3759 used to diagnose liver cancer.
XX
XX KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatocytic;
XX KW metastatic liver tumour; cytotoxic; expression profile; disease state;
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX
XX EN W0200229103-A2.
XX
XX PD 11-Apr-2002.
XX
XX PF 02-OCT-2001; 2001WO-US30589.
XX
XX PR 02-OCT-2000; 2000US-237054P.
XX
XX PA (GENE-) GENE LOGIC INC.
XX
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
XX WP1; 2002-426119/45.
XX
XX DR Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX PT liver tissue sample -
XX
XX CC Claim 1; SEQ ID NO 3759; 298bp; English.
XX
XX PS The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatocytic, and
XX CC cytosaric activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 2799 BP; 792 A; 605 C; 569 G; 833 T; 0 other;

Query Match 74.3%; Score 1225.2; DB 24; Length 2799;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 26 TGCAATCAGATGCTATGAATGAGACTTCACTTCTCTCTGATACAGCTGATGTTAC 85
Db 2 TGCAAAAGATGCTGTAATGAGACTACAG---TTCTGCTGATACAACTCAATGTTTAC 58
Qy 86 TTTAGCTCTGGAGTTGTGAAAAGGTCGTGTGTGTCGCCCAAGAAATTCAGCCACTGATG 145
Db 59 TTTAGCTCTGGAGTTGTGAAAAGGTCGTGTGTGTGTCGCCCAAGAAATTCAGCCCTTGGATG 118
Qy 146 AATTAAGAAATCTCGATGAACTGTGTCAGAGAGGATGATGATGATGATGCA 205

```



```

Db 119 AATATGAAGACATCTGAAAGAACTGTTCAGAGAGGTCAATGAGGTGACTGTACATGCA 178
Qy 206 TCTTCAGCTTCATTTCTTTGATGCCCAAGCCCATCTACTCTTAATTTGAAGTTAT 265
Db 179 TCTTCAGCTTCATTTCTTTGATGCCCAAGCCCATCTACTCTTAATTTGAAGTTAT 238
Qy 266 CCTGATCTTTAACTAAAGTGAAGTATATCAAGAGCTGTTAAGAGATG 325
Db 239 CTTACATCTTTAACTAAAGTGAAGTATATCAAGAGCTGTTAAGAGATG 298
Qy 326 GCAGAACTTCAAAAAGACATTTTGTCTATATTTTCAAGATCAAGAAATCATGTG 385
Db 299 TCAGAAATTTCAAAAAGATATTTTGTCTATATTTTCAAGAAATCATGTG 358
Qy 386 ACATTTATGACATCTTAAAGAGTTCTGAAGATATATGTTTCAAAATGAAGTTATG 445
Db 359 GCAATTTATGACATCTTAAAGAGTTCTGAAGATATATGTTTCAAAATGAAGTTATG 418
Qy 446 AAGAACTTCAAGAGATGATGTTGTCTTGAAGATGCTGTTTCCCTTTGCT 505
Db 419 AAAAACTTCAAGAGATGATGTTGTCTTGAAGATGCTGTTTCCCTTTGCT 478
Qy 506 GAGCTGTGCGCGAGATCTTAAATACCTTTGTCTACAGCTTCCGCTTCTCTGCG 565
Db 479 GAGCTGTGCGCGAGATCTTAAATACCTTTGTCTACAGCTTCCGCTTCTCTGCG 538
Qy 566 TACGCAATTTGAAAGCATATGAGAGAGCTCTGTTTCCCTTCTTATGCTGTGCT 625
Db 539 TACTCATTTTGAAGGACAGTGAAGATTTATTTTCCCTTCTTACGATCTGTGCT 598
Qy 626 ATGTGAGAACTAAGTGAAGCAATGACTTTCATAGAGAGGTTAAAAATATGATCTATG 685
Db 599 ATGTGAGAACTAAGTGAAGCAATGACTTTCATAGAGAGGTTAAAAATATGATCTATG 658
Qy 686 CTTTATTTTGAATTTTGTTCCTCAATTTTGAATTTTGAATTTTGAATTTTGAAT 745
Db 659 CTTTATTTTGAATTTTGTTCCTCAATTTTGAATTTTGAATTTTGAATTTTGAAT 718
Qy 746 GAACTTCAAGAAAGCCCACTTCTTATCTGAAGCAATGAGCAAAAGCTGATGCTT 805
Db 719 GAACTTCAAGAAAGCCCACTTCTTATCTGAAGCAATGAGCAAAAGCTGATGCTT 778
Qy 806 ATTCGAACTTCTGAGATTTTCAATTTTCTTCAAGCTTCTTCAAGATGTTGCTT 865
Db 779 ATTCGAACTTCTGAGATTTTCAATTTTCTTCAAGCTTCTTCAAGATGTTGCTT 838
Qy 866 GAGAGACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 925
Db 839 GAGAGACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 898
Qy 926 AGCTTCTGAGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 985
Db 899 AGCTTCTGAGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 958
Qy 986 GAAAGAAAGGCGCAATGATATGATGATGATGATGATGATGATGATGATGATGATG 1045
Db 959 GAAAGAAAGGCGCAATGATATGATGATGATGATGATGATGATGATGATGATGATG 1018
Qy 1046 AGATTTGATGAGATTAACCAAGATCTTATGAGACTCAATATCTGAGCTGTACAGTGATA 1105
Db 1019 AGATTTGATGAGATTAACCAAGATCTTATGAGACTCAATATCTGAGCTGTACAGTGATA 1078
Qy 1106 CCCGCAATGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1165
Db 1079 CCCGCAATGATCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1138
Qy 1166 AATGAGATTAAGAGCTTATTAACAATGAGGCTTATGAGGCTTATGAGGCTTATGAG 1225
Db 1139 AATGAGATTAAGAGCTTATTAACAATGAGGCTTATGAGGCTTATGAGGCTTATGAG 1198
Qy 1226 GATCAGCTTGAATACATAGCTCAATGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA 1285

```

```

Db 1199 GATCAACCTGATATATATTGCTCATGATGAGGCCAAGGAGACGCTGTAGAGTGACTTC 1258
Qy 1286 AAAAATATGACAGCGAAGATTTACTGAGGCGTTTGAGAACAGATCATTAACGATTCCT 1345
Db 1259 AACCAATGTCAGATACAGACTGCTGATGATGATGATGATGATGATGATGATGATGAT 1318
Qy 1346 TATTAAGAAATGCTATGATGATTTATCAAGAAATTTCAACATGATCAACCTGTAAGCCCTTA 1405
Db 1319 TATTAAGAAATGCTATGATGATTTATCAAGAAATTTCAACATGATCAACCTGTAAGCCCTTG 1378
Qy 1406 GATGAGACATCTTCTGATGATGATTTGTATGATGATGATGATGATGATGATGATGATGAT 1465
Db 1379 GATGAGACATCTTCTGATGATTTGTATGATGATGATGATGATGATGATGATGATGAT 1438
Qy 1466 TCACTGCGCCATGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1525
Db 1439 GTTGAGGCCCAACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1498
Qy 1526 CTGACCTGTGTGCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1585
Db 1499 CTGCTGTGTGTGCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1558
Qy 1586 AAATTTAATTAATTAAGATGAAAGGAAATGATGATGATGATGATGATGATGATGATGAT 1631
Db 1559 AAGTTGCTGTAAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 1604

RESULT 13
AAB06844
ID AAB06844 standard; DNA, 1713 BP.
XX
AC AAB06844;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human breast cancer specific gene-2 (BCSG-2) #1.
XX
KW Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
XX breast cancer; therapeutic; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN M020013779-A2.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000MO-US32056.
XX
PR 23-NOV-1999; 99US-0166973.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Caferkey R, Recipon H, Sun Y;
XX
DR WPI; 2001-367602/38.
XX
PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
XX imaging, preventing and treating cancers, particularly breast cancer -
XX
PS Claim 1; Page 49; 66pp; English.
XX
CC The invention relates human breast cancer specific genes (BCSG's) and
XX their corresponding proteins. BCSG is useful for diagnosing, staging,
XX monitoring, imaging, preventing and treating breast cancers. BCSG is also
XX useful for inducing an immune response against a target cell expressing
XX BCSG. The invention also provide methods for detecting genetic lesions or
XX mutations in BCSG, thereby determining if a human with the genetic lesion
XX is at risk for breast cancer or has breast cancer. BCSG antibodies
XX labelled with paramagnetic ions or radioisotopes is useful for imaging
XX breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is
XX useful for treating breast cancer. BCSG is useful in the rational design
XX of new therapeutics for imaging and treating cancers. BCSG is also used
XX in gene therapy. The present DNA sequence is human breast cancer specific

```


DR P-PSDB: ABG05523.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 5514; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantifying a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1859 BP; 570 A; 337 C; 397 G; 555 T; 0 other;
Query Match 73.2%; Score 1207; DB 23; Length 1859;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1386; Conservative 0; Mismatches 225; Indels 4; Gaps 4;
QY 21 TGCATTGATGATGAGATGCTATGAATGAGACTGAGCTTCTCCTGATACAGTGAAGT 80
DB 1 TGCATTGATGATGAGATGCTATGAATGAGACTGAGCTTCTCCTGATACAGTGAAGT 60
QY 81 GTTACTTATGCTCTGGAGATTGTGGAAGGTGCTGGTGGCCACAGAAATTCAGCCACT 140
DB 61 TTTGCTTTAGCTCTGGAGATTGTGGAAGGTGCTGGTGGCCACAGAAATTCAGCCACT 120
QY 141 GGATGAATATGAAGCAATCTGATGAACTTGTCCAGAGAGCTATGAGTGAAT 200
DB 121 GGATGAATATGAAGCAATCTGATGAACTTGTCCAGAGAGCTATGAGTGAAT 180
QY 201 TGGCATCTTCAGCTCCATTTCTTCCATCCCAAGCCATCTCTTAAATTTGAAG 260
DB 181 TGGCATCTTCAGCTCCATTTCTTCCATCCCAAGCCATCTCTTAAATTTGAAG 240
QY 261 TTTATCTGATATCTTTAACTAAAGTGTGAGATTTATCAAGCAGCTGTAAAGA 320
DB 241 TTTATCTGATATCTTTAACTAAAGTGTGAGATTTATCAAGCAGCTGTAAAGA 300
QY 321 GATGGCAGAACTTCCAAAAGACATTTTGTCTATTTTTCACAGTACAGAAATCA 380
DB 301 GATGGCAGAACTTCCAAAAGACATTTTGTCTATTTTTCACAGTACAGAAATCA 360
QY 381 TGTGACATTTTATGACATCTTAAAGTCTGTAAAGATATAGTTCAATTAAGAAG 440
DB 361 TGTGACATTTTATGACATCTTAAAGTCTGTAAAGATATAGTTCAATTAAGAAG 420
QY 441 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGACAGTGTGTTCCCT 500
DB 421 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGACAGTGTGTTCCCT 480
QY 501 TTGTGAGCTGTGGCCGAGTATCTTAAATACCTTTTGTACAGCCTCCGCTTCTCTC 560
DB 481 GTATGAGCTGTGGCCGAGTATCTTAAATACCTTTTGTACAGCCTCCGCTTCTCTC 540

QY 561 CTGGCTACCGAATTTGAAAAGCAATAGTGAAGACTTGTCTTCTCTTATGTGCTTG 620
DB 541 CTGGCTACCTTTTGAAGAAACATAGTGAAGATTTATTTTCTCTCTTCTTACTACT 600
QY 621 TTGTATGATCAGAACTAAGTGAAGCAATGATCTTCTTATGAGAGGTAAATATGATCT 680
DB 601 TTGTATGATCAGAACTAAGTGAAGCAATGATCTTCTTATGAGAGGTAAATATGATCT 660
QY 681 ATGTGCTTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGATGGAGTCACTTCT 740
DB 661 ATGTGCTTTATTTGATCTTTTGGTCCAAATATTTGACATGAAGATGGAGTCACTTCT 720
QY 741 ACAGTGAAGTTCTAGGAAGCCCACTAAGTATCTGACATGCAAAAGCTGACATAT 800
DB 721 ATAGTGAAGTTCTAGGAAGCCCACTAAGTATCTGACATGCAAAAGCTGACATAT 780
QY 801 GGCTTATTCGAAACTACTGGGATTTTCAATTTTCTCCACCTTTTCCAAATGTTAGT 860
DB 781 GGCTTATTCGAAACTACTGGGATTTTCAATTTTCTCCACCTTTTCCAAATGTTAGT 840
QY 861 TCGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGTGTG 920
DB 841 TTGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGTGTG 900
QY 921 TCCAGACCTCTGAGAAATGTTGTGTTGTTTCTCTGGGCTCATGCTCAGTAA 980
DB 901 TACAGACCTCTGAGAAATGTTGTGTTGTTTCTCTGGGCTCATGCTCAGTAA 960
QY 981 CGTCAAGAAAGGCGCAATGTAATTCATCAAGCCCTTCCAAATCCCAAAAGGTTCC 1040
DB 961 TGACAGAAAGGCGCAATGTAATTCATCAAGCCCTTCCAAATCCCAAAAGGTTCC 1020
QY 1041 TGT-GGAGATTTGAT-GGGAATTAACAGATTAATCTTGAAGTCAATCTCGGCTGTA 1098
DB 1021 TGTGGAGATTTGATGGGATTTAAACAGATTAATCTTGAAGTCAATCTCGGCTGTA 1080
QY 1099 GT-GGATACCCCAAG-TGATCTTCTGTGTCATCCAAACCAAAAGCTTTTACCTCAT 1156
DB 1081 GTGGAGATTCACCAATGATGACCTTCTAGTCACTCAAAAGCAAGCTTTTATTAATCAT 1140
QY 1157 GGTGAATGAATGGGATCTATGAAGTATTTTCAATGGGCTCTTATGAGTGGAGTTCCC 1216
DB 1141 GGTGAAGCAATGGGATCTATGAAGTATTTTCAATGGGCTCTTATGAGTGGAGTTCCA 1200
QY 1217 ATATTGGTATGAGCTGATTAATGATAGTCAATGAGGCTTGAAGAGAGCTGTAGAA 1276
DB 1201 TTGTTGCGATCAACCTGATTAATGATAGTCAATGAGGCTTGAAGAGAGCTGTAGAA 1260
QY 1277 ATAACTTCAAACTATGACAAAGGAAATTTACTGAAGGCTTGAAGAGAGCTGTAGAA 1336
DB 1261 GTGACCTTCAACCAATGATGACAAAGGAAATTTACTGAAGGCTTGAAGAGAGCTGTAGAA 1320
QY 1337 GATTCCTTTTAAAGAGATGCTATGAGATTTTCAAGATTTTCAACGATGACACCTGTA 1396
DB 1321 GATTCCTTTTAAAGAGATGCTATGAGATTTTCAAGATTTTCAACGATGACACCTGTA 1380
QY 1397 AAGCCCTGATGAGAGAGCTCTTCTGATGAGTGTTCATGCGCCCAAAAGAGAGCTAA 1456
DB 1381 AAGCCCTGATGAGAGAGCTCTTCTGATGAGTGTTCATGCGCCCAAAAGAGAGCTAA 1440
QY 1457 CACCTGATGAGAGCTGCGCATGACCTCACTGTTTCCAGACTCTTGAATGATGAT 1516
DB 1441 CACCTGATGAGAGCTGCGCATGACCTCACTGTTTCCAGACTCTTGAATGATGAT 1500
QY 1517 GGGTTCCTGCTGACCTGTGTGGAAGCTGCTATATTTCTTGTTCACAAAATGTTTATAT 1576
DB 1501 GGGTTCCTGCTGCTGCTGTGTGGAAGCTGCTATATTTCTGTCACAAAATGTTTCTGTT 1560
QY 1577 TCTCTCAAAATTTTAAATTAAGTAAAGATGAGGAGATAGATCTTCT 1631
DB 1561 TCTCTCAAAATTTTAAATTAAGTAAAGATGAGGAGATAGATCTTCTATC 1615

RESULT 15
AAS69711
ID AAS69711 standard; cDNA; 2802 BP.
XX
AC AAS69711;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #515.
XX
KM Human, chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Dermanac RT, Liu C, Tang YT;
XX
DR WPI, 2001-639362/73.
DR P-PSDB; ABG05524.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 5515; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 2802 BP; 791 A; 605 C; 571 G; 835 T; 0 other;

Query Match 72.7%; Score 1199.4; DB 23; Length 2802;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1367; Conservative 0; Mismatches 236; Indels 6; Gaps 2;

QY 26 TGCATCGAGATGTCATGAATGAGCTTCAGCTCTTCCTCGATACAGCTGAGCTGTAC 85
DB 2 TGCACAGAGATGCTCTGAATGAGCTACAG--TTCTGCTGATACACATCAGTTTATAC 58

QY 86 TTATGCTCTGGAGCTTGAGAAAGGTGTGTGTGTGCGCCACAGAAATTCAGCCACTGATG 145
DB 59 TTATGCTCTGGAGCTTGAGAAAGGTGTGTGTGTGCGCCACAGAAATTCAGCCACTTGTGATG 118

QY 146 AATATAAGACAATCTCGATGAACTTGTCCAGAGAGTCATGAGGTGATGTTGCA 205
DB 119 AATATGAAGACAATCTCGAAGAACTTGTTCAGAGAGTCATGAGGTGATGTTGCA 178

QY 206 TCTTCAGCTTCCATTTCTTTCGATCCCAAGGCCATCTACTTAAATTTGAAGTTAT 265
DB 179 TCTTCAGCTTCCATTTCTTTCGATCCCAAGGCCATCTACTTAAATTTGAAGTTAT 238

QY 266 CCTGTATCTTTAACTAAACTGAGTTTGAGATATTTCAAGCAAGCTGTTAGAGATG 325
DB 239 CTTACATCTTTAACTAAACTGAGTTTGAGATATTTCAAGCAAGCTGTTAGAGATG 298

QY 326 GCAGAACTTCCAAAGACACATTTGGTCAATTTTTCACAGATGCAAGAAATCATGTG 385
DB 299 TCAGAAATTCAAAGAAAGTACATTTTGGTTACTTTTTCACAGAAAGAAATCTGTG 358

QY 386 ACATTTAATGACATCTTAGAAAAGTTCTGAAGATATAGTTTCAATTAAGAACTTAG 445
DB 359 GCATTTAATGACATCTTAGAAAAGTTCTGAAGATATAGTTTCAATTAAGAACTTAG 418

QY 446 AAGAACTACAGAGATCAAGATTTGATGTTGTTCTTCAAGATGCTGTTTCCCTTGT 505
DB 419 AAAAACTACAGAGATCAAGATTTGATGTTGTTCTTCAAGATGCTGTTTCCCTTGT 478

QY 506 GAGCTGCTGGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTCTGCTGCG 565
DB 479 GAGCTGCTGGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTCTGCTGCG 538

QY 566 TACGCAATTAAGAAAGCATAGTGAAGAGCTTCTGTTCCCTCTCTTATGTCCTGTTGT 625
DB 539 TACTCATTTGAAGAGCAGCTGAGATTTATTTCCCTCTCTTATGTCCTGTTGT 598

QY 626 ATGTGAGACTAAGTGAACAAATGACTTTCTATGAGAGGTTAAATAATGATCTATGT 685
DB 599 ATGTGAGACTAAGTGAACAAATGACTTTCTATGAGAGGTTAAATAATGATCTATGT 658

QY 686 CTTATTTTGAATTTGGTTCCTCAATTTTGAATGAGAAAGTGGATGCTGCTACAGT 745
DB 659 CTTATTTTGAATTTGGTTCCTCAATTTTGAATGAGAAAGTGGATGCTGCTACAGT 718

QY 746 GAAGTTCTAGGAAGACCCCTAGCTTATCTGAGACATGAGCAAAAGCTGACATATGCTT 805
DB 719 GAAGTTCTAGGAAGACCCCTAGCTTATCTGAGACATGAGCAAAAGCTGACATATGCTT 778

QY 806 ATTGGAATCTAGGAGATTTTCAATTTCTACCCACTTTACCAATGTTGAGTTGCTT 865
DB 779 ATCGGAATCTAGGAGATTTTCAATTTCTACCCACTTTACCAATGTTGAGTTGCTT 838

QY 866 GAGGACTCCACCTGCAAACTGCAAAACCTTACCCAGAAAGAAATGGAAGTTGTCAG 925
DB 839 GAGGACTCCACCTGCAAACTGCAAAACCTTACCCAGAAAGAAATGGAAGTTGTCAG 898

QY 926 AGCTCTGGAAGAAATGTTGTGTGTGTTTCTCTGAGGTCAGATGTCATACACCTCA 985
DB 899 AGCTCTGGAAGAAATGTTGTGTGTGTTTCTCTGAGGTCAGATGTCATACACCTCA 958

QY 986 GAAAGAAAGGCAATGATTTGATGATGAGCTTCTGCAAGATCCCAAAAGTTCTGTG- 1044
DB 959 GAAAGAAAGGCAATGATTTGATGATGAGCTTCTGCAAGATCCCAAAAGTTCTGTG- 1018

QY 1045 --GAGATTTGATGAGAAATTAACAGATACTTTAGAGTCAATCTGGGCTGTCAAGTGG 1102
DB 1019 GAGATTTGATGAGAAATTAACAGATACTTTAGAGTCAATCTGGGCTGTCAAGTGG 1078

QY 1103 ATACCCAGAAATATCTTCTTGTCTATCCCAAAACCAAGCTTTTATCACTACGTGGA 1162
DB 1079 ATACCCAGAAATATCTTCTTGTCTATCCCAAAACCAAGCTTTTATCACTACGTGGA 1138

QY 1163 ATGAATGGATCTATGAAGCTATTTACATGAGGCTCTATGTTGAGAGTTCCCATATTT 1222
DB 1139 GCAATGAGATCTATGAAGCTATTTACATGAGGCTCTATGTTGAGAGTTCCCATATTT 1198

Oy 1223 GGATCAGCTTGTATACATAGCTCACATGAAAGCCAAAGAGAGCAGCTGTAGAATAAAC 1282
Db 1199 TTGATCAACCTGATTAATATGCTCACATGAAAGCCAAAGAGAGCAGCTGTAGAGTGAC 1258
Oy 1283 TTCAAACTATGACAAGGAGATTACTGAGGGCTTTGAGACAGTCATTAACGATTCC 1342
Db 1259 TTCAAACAATGTGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCT 1318
Oy 1343 TCTTATAAGGAATGCTATGAGATTATCAAGATTTCACCATGATCAACCTGTAAAGCCC 1402
Db 1319 TCATATATAAGGAATATTATGAAATTAACAAGAAATTCACAAATGATCAACAGTGAAGCCC 1378
Oy 1403 CTGATCGAGCAGCTTCTGATCGATCGATTGTCAATGCCCAAAAGAGCCAGCACCTG 1462
Db 1379 CTGATCGAGCAGCTTCTGATCGATCGATTGTCAATGCCCAAAAGAGCCAGCACCTT 1438
Oy 1463 CGATCAGCTGCCATGATGACCTGCTGTTCCAGCAGTAATCTATAGATGATGGGTTG 1522
Db 1439 CGAGTGGAGCCCAACACCTCACCCTGTTCCAGTACCACTCTTGATGATGATGGGTTG 1498
Oy 1523 CTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAATGTTTTTATTTCCCTGT 1582
Db 1499 CTGCTGCTGTGTGGCAACCGTGTCTATTATCAATCAAAAGTGTGTGTGTTGTTTC 1558
Oy 1583 CAAAAATTTAATAAATACTAGAAAGATAGAAAAGAGGAAATAGATCTTTC 1631
Db 1559 TGGAACTTGTCTAGAAAAGGAAAGGAAAGGAAAGGAAATAGTATATTC 1607

Search completed: December 7, 2003, 05:47:47
Job time : 355 secs

OY	301	TATCAGCAGCTGGTTAAGAGTGGGCGAACTTCCAAAAGACATTTGGTCATATT	360
Db	304	TATCAGCAGCTGGTTAAGAGTGGGCGAACTTCCAAAAGACATTTGGTCATATT	363
OY	361	TTCAACAATACAGAANAATCATGTGGACATTTAATGACATACTTGAAGATTCTGTAAAGA	420
Db	364	TTCAACAATACAGAANAATCATGTGGACATTTAATGACATACTTGAAGATTCTGTAAAGA	423
OY	421	TATAGTTCCAAATAAAGAACTTATGAGAAACTACAGAGATCAGATTTGATGTTGTCT	480
Db	424	TATAGTTCCAAATAAAGAACTTATGAGAAACTACAGAGATCAGATTTGATGTTGTCT	483
OY	481	TGCAGATCTGTTTTCCCTTTGGAGCTGCTGGCCGAGTATCTAAATAACCTTTGT	540
Db	484	TGCAGATCTGTTTTCCCTTTGGAGCTGCTGGCCGAGTATCTAAATAACCTTTGT	543
OY	541	CTACAGCTCGGCTTCTCTCTGGCTACGCAATTGMAAAGATAGTGGAGACTTCTGT	600
Db	544	CTACAGCTCGGCTTCTCTCTGGCTACGCAATTGMAAAGATAGTGGAGACTTCTGT	603
OY	601	CCCTCTTCTATGTGCTGCTGTTGTTATGTCAGAATTAAGTACCAATGACTTTCATAGA	660
Db	604	CCCTCTTCTATGTGCTGCTGTTGTTATGTCAGAATTAAGTACCAATGACTTTCATAGA	663
OY	661	GAGGGTAAAAAATATGATCTATGCTTATTTTGAATTTGGTCCAAATATTTGAAT	720
Db	664	GAGGGTAAAAAATATGATCTATGCTTATTTTGAATTTGGTCCAAATATTTGAAT	723
OY	721	GAGAGATGGGATCAGTTCTACAGTGAAGTTCTAGAGAGACCCTACGTTATCTGAGAC	780
Db	724	GAGAGATGGGATCAGTTCTACAGTGAAGTTCTAGAGAGACCCTACGTTATCTGAGAC	783
OY	781	AATGGCAAAAGCTGACATATGGCTTATTCGAAACTACTGGGATTTTCAATTTCTCACC	840
Db	784	AATGGCAAAAGCTGACATATGGCTTATTCGAAACTACTGGGATTTTCAATTTCTCACC	843
OY	841	ACTCTTACCAATGTGAGTTCGTTGGAGAGCTCCACTGCAAACTGGCAAACTCCCTAC	900
Db	844	ACTCTTACCAATGTGAGTTCGTTGGAGAGCTCCACTGCAAACTGGCAAACTCCCTAC	903
OY	901	GAAGGAAATGGAAGATTGTTCACAGCTCTGGAGAAATAGGTGTTGTGTTCTCT	960
Db	904	GAAGGAAATGGAAGATTGTTCACAGCTCTGGAGAAATAGGTGTTGTGTTCTCT	963
OY	961	GGGGTCGATGTCAGTAAACGTCAGAGAAAGGGCCAAATGTAATTGCATCAGCCCTTGC	1020
Db	964	GGGGTCGATGTCAGTAAACGTCAGAGAAAGGGCCAAATGTAATTGCATCAGCCCTTGC	1023
OY	1021	CAGATATCCCAAAAGGTTCTGTGGAGATTTGATGGGAATTAACCAAGATCTTTAGAGCT	1080
Db	1024	CAGATATCCCAAAAGGTTCTGTGGAGATTTGATGGGAATTAACCAAGATCTTTAGAGCT	1083
OY	1081	CAATACTGGGCTGTCAAGTGAATCCCGAGATATCTTCTGTGTCATCCCAAAACAA	1140
Db	1084	CAATACTGGGCTGTCAAGTGAATCCCGAGATATCTTCTGTGTCATCCCAAAACAA	1143
OY	1141	AGCTTTTATCACTCATGTGTGATGAATGGAATCTATGAAGCTATTTACATGGGGTCC	1200
Db	1144	AGCTTTTATCACTCATGTGTGATGAATGGAATCTATGAAGCTATTTACATGGGGTCC	1203
OY	1201	TATGGTGGAGATTTCCCATATTTGTGTGATCAGTTGATTAACATAGTCACTGAAAGCCAA	1260
Db	1204	TATGGTGGAGATTTCCCATATTTGTGTGATCAGTTGATTAACATAGTCACTGAAAGCCAA	1263
OY	1261	AGGACAGCTGTGAGAAATTTAACTTCAAAACTATGCAAGCCAGAAATTTATGAGGGCTTT	1320
Db	1264	AGGACAGCTGTGAGAAATTTAACTTCAAAACTATGCAAGCTTATGAGGGCTTTATGACT	1323
OY	1321	GAGAACAGTCATTACCGATTCCTCTTATTAAGAGATGTATAGATTTATCAAGATTTCA	1380
Db	1324	GAGAACAGTAAATTAATGATCCTTTATTAATGAAGATGTATGAATTTATCAAGATTTCA	1383
OY	1381	CCATGATCAACTGTAAAGCCCTTATGATGACAGCTCTTCTGATCGAGTTTGTATGCG	1440

```
Db      1384 TCATGATCAACAGTGAAGGCCCTTGATCGAGCAGTCCTTGTGGATTGTAATTTGTATGGG 1443
QY      1441 CCACAAAGAGCCAAACACTGCGCATCAGCTGCCATGACTCACCTGGTTGCAGACTA 1500
Db      1444 CCAATAAGAGACCAGAACAACCTTCGGGTTGACAGCCCAACAGCACTCAGCTGGTTCCAGTACA 1503
QY      1501 CTCTATATATGTGANTGGTGGTCCGCTGACACCTGCGGCAACCTGCTATATTCTGTCTAC 1560
Db      1504 CTCCTTGATGTGACTGGGTTCTGCTGCTGCGCTGTGTGGCACTGTGATATTATCATCTAC 1563
QY      1561 AAAATGTTTTTATTTATTTCTGTCAAAAATTTAATAAACCTAGAAAAGTAGAAAAGAGGA 1620
Db      1564 AAATGCTGTGTTTGTGTCTG---GAGTTGTGTGAACAGAAAAGAGGAAAAGAGA 1620
QY      1621 ATGATCTTTC 1631
Db      1621 TTAATTACGTC 1631

RESULT 2
US-09-356-806-39
; Sequence 39, Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvan, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucanosephyltransferase 2B4 (UGR2B4), 2B7 (UGR2B7) and
; TITLE OF INVENTION: 2B15 (UGR2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/09/356,806
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)...(1584)
; US-09-356-806-39

Query Match          75.6%; Score 1247.8; DB 4; Length 1854;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1884; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
```

```

Qy 381 TGTGACATTTAATGACATCTTAGAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 440
Db 361 TGTCAATATTTGGAGACATTAAGTAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 420
Qy 441 TTAATGAAGAACTACAGAGTCAAGATTTGATGTTTCTTGACAGATGCTGTTTCCCT 500
Db 421 TTAATGAAGAAAGTCAAGAGTCAAGATTTGATGTTTCTTGACAGATGCTGTTTCCCT 480
Qy 501 TTGGTGAAGCTGCTGGCCGAGTACTTAATAATCCCTTTGTCTAAGGCTCCGCTTCTC 560
Db 481 GTAGTGAAGCTGCTGGCCGAGTACTTAATAATCCCTTTGTGTACAGTCAAGCTTCTC 540
Qy 561 CTGGCTACGCAATTTGAAAGCATAGTGAAGGACTTCTGTTCCCTCCCTCATGCTGCTG 620
Db 541 CTGGCTACGCAATTTGAAAGCATAGTGAAGGACTTATTTTCCCTCCCTCATGCTGCTG 600
Qy 621 TTGTTATGTGAGAACTAAGTACCAATGACTTTTCAATAGAGGATTAATAATATGATCT 680
Db 601 TTGTTATGTGAGAACTAAGTACCAATGACTTTTCAATAGAGGATTAATAATATGATCT 660
Qy 681 ATGCTCTTATTTTGAATTTTGTGTTCCAAATTTTGTGATGAAGAGTGGGATCAATCT 740
Db 661 ATGCTCTTATTTTGAATTTTGTGTTCCAAATTTTGTGATGAAGAGTGGGATCAATCT 720
Qy 741 ACAGTGAAGTCTAGAAAGCACTACGTTATCTGAGACAAATGSCAAAGCTGACATAT 800
Db 721 ATAGTGAAGTCTAGAAAGCACTACGTTATCTGAGACAAATGSCAAAGCTGACATAT 780
Qy 801 GGCTTATTTGAACTACTGAGATTTTCAATTTCTTCAACCACTCTTACCAATGTTAGT 860
Db 781 GGCTTATTTGAACTACTGAGATTTTCAATTTCTTCAACCACTCTTACCAATGTTAGT 840
Qy 861 TCGTTGAGAGTCTCACTGCAACCTGCAAACTGCAACCAAGAGAAATGGAAGATTG 920
Db 841 TCGTTGAGAGTCTCACTGCAACCTGCAAACTGCAACCAAGAGAAATGGAAGATTG 900
Qy 921 TCCAGAGCTCTGAGAAATGATGTTGTGTTTCTGTTGTTTCTGTTGTTGTTGTTGTTG 980
Db 901 TACAGAGCTCTGAGAAATGATGTTGTGTTTCTGTTGTTTCTGTTGTTGTTGTTGTTG 960
Qy 981 CGTCAAGAAAGGCGCAATGTAATGATCAGCCCTTCCCAAGATCCCAAAAGGTTT 1040
Db 961 TGAAGAAAGGCGCAATGTAATGATCAGCCCTTCCCAAGATCCCAAAAGGTTT 1020
Qy 1041 TGTGAGATTTGATGGAATTAACCAAGATCTTGAAGCTCAATGCTGCTGTAAGT 1100
Db 1021 TGTGAGATTTGATGGAATTAACCAAGATCTTGAAGCTCAATGCTGCTGTAAGT 1080
Qy 1101 GGATACCCCAAGATGATCTTCTTGTGATCCCAAAAGCTTTTATCACTGATGTTG 1160
Db 1081 GGATACCCCAAGATGATCTTCTTGTGATCCCAAAAGCTTTTATCACTGATGTTG 1140
Qy 1161 GAATGAATGGATCTATGAAGCTATTTACATGAGGTTCCCTATGTTGAGATTTCCATAT 1220
Db 1141 GAGCAATGGATCTATGAAGCTATTTACATGAGGTTCCCTATGTTGAGATTTCCATAT 1200
Qy 1221 TTGGTGAATGAGTTGATTAACATGCTCAATGAAGGCGCAAGAGAGCTGTAGAAATTA 1280
Db 1201 TTGGTGAATGAGTTGATTAACATGCTCAATGAAGGCGCAAGAGAGCTGTAGAAATTA 1260
Qy 1281 ACTTCAAAATCTATGAAGCAAGATTTTACGAGGCTTTGAGAAAGCTCAATTCATAT 1340
Db 1261 ACTTCAAAATCTATGAAGCAAGATTTTACGAGGCTTTGAGAAAGCTCAATTCATAT 1320
Qy 1341 CCTCTTAAAGAGATGCTATGAGATTTCAAGATTTCAATGATCAATCTGTAAGC 1400
Db 1321 CCTCTTAAAGAGATGCTATGAGATTTCAAGATTTCAATGATCAATCTGTAAGC 1380
Qy 1401 CCTGATGAGAGCTCTTGTGATGAGTTGTGATGCGCAAGAGAGCAAGCAACC 1460
Db 1381 CCTGATGAGAGCTCTTGTGATGAGTTGTGATGCGCAAGAGAGCAAGCAACC 1440

```

```

Qy 1461 TGGATGAGCTGCCATGACCTGATGCTGACAGACTATAGATGATGAGGCT 1520
Db 1441 TTGGGTTGAGCGCCAGACCTGATGCTGATGATGATGATGATGATGATGATGATGATGAT 1500
Qy 1521 TCTGCTGACCTGTTGAGCACTGCTATATTTCTTTCACAAATGTTTATTTTCT 1580
Db 1501 TCTGCTGAGCTGTTGAGCACTGCTATATTTCTTTCACAAATGTTTATTTTCT 1560
Qy 1581 GTCAAAATTTAATTAAGATGAGAAAGTAAAGAGGGAATGATCTTTC 1631
Db 1561 TCTGAAATTTGCTAAGAAAGCAAGAGGAAATTAATGATTAATATC 1611

RESULT 3
US-09-180-852-1
Sequence 1, Application US/09180852
Patent No. 6287834
GENERAL INFORMATION:
APPLICANT: BELANGER, Alain
APPLICANT: HUM, Dean W.
APPLICANT: BEAULIEU, Martin
TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
FILE REFERENCE: 1259-449
CURRENT APPLICATION NUMBER: US/09/180,852
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: PCT/CA97/00328
EARLIER FILING DATE: 1997-05-16
EARLIER APPLICATION NUMBER: US 08/649,319
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 2107
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (52)..(1644)
US-09-180-852-1

Query Match 70.3%; Score 1159.6; DB 3; Length 2107;
Best Local Similarity 82.2%; Pred. No. 0;
Matches 1346; Conservative 0; Mismatches 289; Indels 3; Gaps 1;

Qy 1 AGCACTGAAAAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 18 AACAACTGAAAAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 77
Qy 61 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
Db 78 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 137
Qy 121 GCCCAGAAATGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 138 GCCCAGAAATGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 197
Qy 181 AGGTGATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
Db 198 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 257
Qy 241 ATCTACTTAAATTTGAAGTTTATCTGATCTTTTAACTAAATGATGATGATGATGAT 300
Db 258 ATCTACTTAAATTTGAAGTTTATCTGATCTTTTAACTAAATGATGATGATGATGAT 317
Qy 301 TATCAAGCAGCTGTTAAGAGTGGGAGAG---ACTTCAAAAGACATTTTGTGATA 357
Db 318 TTTTATGAATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377
Qy 358 TTTTTCACAGATCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
Db 378 TTTTTCACAGATCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437

```



```

Oy 418 GGATATAGTTTCAATTAAGAACTTATGAGAACTACAGAGTCAAGATTGATGTTGT 477
Db 438 AGATGACGTTTGAACAAAGAACTTATGAGAACTACAGAGTCAAAATTTGATGTTCT 497
Oy 478 TCTTGACAGTCTGTTTTTCCCTTTGATGAGCTGCGCCGAGTACTTAAATACCTT 537
Db 498 TCTGACAGATGCGGTTAATCCCTGTGTGAGCTGCTGAGTCACTTAAATACCTT 557
Oy 538 TGTCTACAGCTCCGCTTCTCTCCGCTGAGCAATTAAGCAATGATGAGGCTTCT 597
Db 558 TCTGTACAGTCTCCGCTTCTCTGTTGCTACACAGTTAGAAAGTGTGAGGATTTCT 617
Oy 598 GTTCCCTCTTCCCTTATGCTGCTGTTGATGTCAGAACTAAGTACCAATGACTTTGAT 657
Db 618 GTTCCCTCTTCCCTTATGATACCTGTTGTTATGTCAGAACTAAGTATCAATTTTCTAT 677
Oy 658 AGAGAGGTTAAATAATGATCTATGCTTATTTGATTTTGGTTCCAAATATTTGA 717
Db 678 GAGAGGATTAATAATGATCTATGATCTTATTTGATTTTGGTTTCAAGCATATGA 737
Oy 718 CATGAAGAAGTGGATCTAGTTCTACAGTGAAGTTCTAGAGAAAGCCACTAGTTATCTGA 777
Db 738 TCTGAAGAAGTGGATCTAGTTCTATGAGAGTTCTAGAGAAAGCCACTAGTTATTTGA 797
Oy 778 GACAAATGGCAAAAGCTGACATATGCTTATTTGAACTAGTGGAGTTTCAATTTCTCA 837
Db 798 GACAAATGGCAAAAGCTGAAATGCTGCTATTTGAACTAGTGGAGTTTGAATTTCTCG 857
Oy 838 CCCACTCTTACCAATGTTGAGTTGTTGAGAGACTGCACTGCAAACTGCCAAACCTT 897
Db 858 CCCACTCTTACCAATGTTGATTTTGTGAGAGACTTCACTGTAACAGCAACCAACCTT 917
Oy 898 ACCGAAGGAATGGAAGAGTTTGTCCAGAGCTCTGGAAGAAATGCTGTTGCTGCTTTC 957
Db 918 GCTTAAGGAAGAGAGAGTTGTCAGAGCTCTGGAAGAAATGCTGTTGCTGCTTTC 977
Oy 958 TCTGGGCTGATGTCAGTAACAGCTCAGAAAGAGGCAATGTAATTTGATCCAGCTT 1017
Db 978 TCTGGGCTGATGTCAGTAACAGCTCAGAAAGAGGCAATGTAATTTGATCCAGCTT 1037
Oy 1018 TGCCAAAGATCCCAAAAGTTCTGTGAGATTTTATGAGAAATTAACAGATCTTTAG 1077
Db 1038 TGCCAAAGATCCCAAAAGTTCTATGAGATTTATGAGAAAGCAAAATTTAGT 1097
Oy 1078 ACTCAATCTGCGCTGTACAGTGAATCCCAAGATGATCTTTGCTGATCCCAAAAC 1137
Db 1098 TTCCAAATCTGCACTGTATAGTGTGTAACCCAGATGACCTTTGCTGATCCCAAAAC 1157
Oy 1138 CAAGCTTTTATCACTCATGTCAGTAAGTGAATGAGATCTATGAAAGCTATTTACATGG 1197
Db 1158 CAAGCTTTTATCACTCATGTCAGTAAGTGAATGAGATCTATGAAAGCTATTTACATGG 1217
Oy 1198 CCTTATGTTGGAGTTCCCATATTTTGTGATCAGCTTGAATACATAGCTCATGTAAG 1257
Db 1218 CCTTATGTTGGAGTTCCCTTTGTTGGGATCAATGATGATGCTCATGTAAG 1277
Oy 1258 CAAGAGAGAGCTGTGAGAAATTAATCTTCAAACTATGACAGCGAAAGTTTACTGAG 1317
Db 1278 CAAGAGAGAGCTGTGAGAAATTAATCTTCAAACTATGACAGCGAAAGTTTACTGAG 1337
Oy 1318 TTTGGAACAGTCAATTAACGATTTCTCTTATTAAGAAAGTGTATGAGATTAACAAG 1377
Db 1338 ATTTGAAGTGTATGATTAATGATCCCTATCTATTAAGAAAGTGTATGAGATTAACAAG 1397
Oy 1378 TCACCATGATCAACCTGTAAAGCCCTTAGATGAGCAGTCTTCTGATGAGTGTGAT 1437
Db 1398 TCATATGATCAACCGGTGAAGCCCTGTGATGAGCAGTCTTCTGATGAGTGTGAT 1457
Oy 1438 GCGCCACAAAGAGCCAAAGCCTGTGATCAGCTGCCATGACCTGCTGATCCAGCA 1497
Db 1458 GCGCCATTAAGAGCCAAAGCCTTTCGGTGTGAGCCCAACCTGATGATCCAGTA 1517

```

```

Oy 1498 CTACTATAGATGATGATGGGTTCTGTCAGCTGTGTGGCACTGCTATATTCTTGT 1557
Db 1518 CCACTCTTTGGATGTGATGATCATTTCTGCTGGCTGTGGCACTATGATTTATGAT 1577
Oy 1558 CACAAATGTTTTTATTTTCTCTGCAAAATTTAATTAAGTGAAGTGAAGAG 1617
Db 1578 CACAAATGTTGCTGTTTGTGTTTCCGAAAGCTTCCGAAAGTGAAGAGAG 1637
Oy 1618 GGAATAGATCTTTCCAA 1635
Db 1638 GGATTAATGATTAACAA 1655

RESULT 4
US-09-356-806-112
; Sequence 112: Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/09/356,806
; CURRENT FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11) ... (1598)
US-09-356-806-112

Query Match 68.3%; Score 1126.8; DB 4; Length 1976;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1315; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

Oy 29 ATCAGATGTTTATGAATGATGATGATGATGATGATGATGATGATGATGATGAT 88
Db 5 ACCAGAGTGTCTGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 64
Oy 89 AGCTTGGAGTTGTGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 148
Db 65 AGCTTGGAGTGTGTGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 124
Oy 149 ATTAAGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 208
Db 125 ATTAAGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
Oy 209 TCAGCTTCATTTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 268
Db 185 TCGGCTTCTACTTTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 244
Oy 269 GTATCTTTAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 328
Db 245 ACATCTTTAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304
Oy 329 ---GACTTCCAAAGACACTTTTGTGATGATGATGATGATGATGATGATGATGAT 385
Db 305 TATGATGTTTCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 364
Oy 386 ACATTTATGACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 445
Db 365 GATTTTATGACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424
Oy 446 AAGAACTACAGAGATCAAGATTTGATGATGATGATGATGATGATGATGATGAT 505

```


425 ATGAACTACAGAGTCAGAACTTGTGATCTCTGCGAGATGCCCTTAATCCCTGTGCT 484
506 GAGCTGTGCGCGAGTTACTTAATAATACCTTGTCTACAGCTCCGCTCTCTCGGC 565
485 GAGCTACTGCGTGAACATTTTAACTATCCCTTCTGTAAGCTCTTCAATCTCTGTGGC 544
566 TACGCAATGGAAGCATAGTGAAGAGCTTCTGTCCCTCTCTCTGTGCTGTGCT 625
545 TACACATTTGAGGAAGATGTGAGAGATTTCTGTTCCTCTCTCTATGATCCTGTGCT 604
626 ATGTCAGACTAATGATGACCAATGATCTTCAATAGAGGGTAAATAATATGATCTATG 685
605 ATGTCAGAAATTAATGATCAATGATTTTCAATGAGAGGATTAATAATATGATCAATG 664
686 CTATATTTGAAATTTGGTCCAAATATTTGACATGAGAGGATGAGTCACTTCAAGT 745
665 CTATATTTGACTTTGGTTCCTCAAAATTTATATGATGAGAGGATGAGTCACTTATAGT 724
746 GAACTTCTAGGAAGACCCACTACGTTATCTGAGCAATGCGAAAGCTGACATATGCTT 805
725 GAACTTCTAGGAAGACCCACTACATTTATGAGCAATGCGAAAGCTGAAATGTGCTC 784
806 ATTCGAACTACTGCGATTTTCAATTTCTGCAACCACTCTTACCAAAATGTTGAGTCTT 865
785 ATTCGAACTACTGCGATTTTGAATTTCTGCGCAATCTTACCAAAATGTTGATTTGTT 844
866 GAGAGACTCCCACTGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTG 925
845 GAGAGACTCCCACTGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTG 904
926 AGCTCTGAGAGAAATGCTGTGTTGCTGTGTTCTGCGGGTGCATGATGATGATGATGAT 985
905 AGCTCTGAGAGAAATGCTGTGTTGCTGTGTTCTGCGGGTGCATGATGATGATGATGAT 964
986 GAAAGAAAGGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1045
965 GAAAGAAAGTGCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1024
1046 AGATTTGATGGGAATTAACCGATATCTTGAAGCTCAATCTGCGCTGTACAGTGTGATA 1105
1025 AGATTTGATGGGAATTAACCGATATCTTGAAGCTCAATCTGCGCTGTACAGTGTGATA 1084
1106 CCCGAGATGATCTTCTGTGCTATCCCAAAACCAAAAGCTTTATCACTCATGATGATGAT 1165
1085 CCCGAGATGATCTTCTGTGCTATCCCAAAACCAAAAGCTTTATCACTCATGATGATGAT 1144
1166 AATGGATCTATGAGGATATTTACATGAGGATGATGATGATGATGATGATGATGATGAT 1225
1145 AATGGATCTATGAGGATATTTACATGAGGATGATGATGATGATGATGATGATGATGAT 1204
1226 GATGAGCTTGTATCAATGCTCATGAGGATGATGAGGATGATGAGGATGATGAGGATGAT 1285
1205 GATGAGCTTGTATCAATGCTCATGAGGATGATGAGGATGATGAGGATGATGAGGATGAT 1264
1286 AAAAATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1345
1265 AGGACCAATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1324
1346 TATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1405
1325 TATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1384
1406 GATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1465
1385 GATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444
1466 TCACTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1525
1445 GTGCGAGCTCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
1526 CTGAGCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1585
1505 CTGAGCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1564

1586 AAATTAATAAACTAGAAATGAGAAAGGAGATAGATCTTCCAAA 1635
1565 AAGCTCCCAAAACAGAAAGAGAAAGAAAGAGATTAATATCAAAA 1614

RESULT 5
US-09-813-918-1
Sequence 1, Application US/09813918
Patent No. 6383789
GENERAL INFORMATION:
APPLICANT: WEBSTER, Marion et al.
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001175
CURRENT APPLICATION NUMBER: US/09/813,918
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1413
TYPE: DNA
ORGANISM: Human
US-09-813-918-1

Query Match 49.6%; Score 818; DB 4; Length 1413;
Best Local Similarity 72.9%; Pred. No. 3e-229;
Matches 1182; Conservative 0; Mismatches 215; Indels 225; Gaps 1;

13 ACAAGATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 72
1 AGAATGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
73 GCTGAGCTGTACTTGTAGCTCTGAGAGTGTGAGAAAGTCTGAGTGTGAGCCACAGAAAT 132
61 TCTCAGTTGTACTTGTAGCTCTGAGAGTGTGAGAAAGTCTGAGTGTGAGCCACAGAAAT 120
133 CAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
121 CAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
193 GACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
181 GACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
253 ATTGGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 312
241 ATTGGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
313 GGTAAAGATGAGGAGAGAACTTCCAAAGACACATTTTGTGATTTTTCACAAAGTACA 372
301 GGTAAAGATGAGGAGAGAACTTCCAAAGACACATTTTGTGATTTTTCACAAAGTACA 360
373 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
361 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
433 TAAAGAACTATGAGAGAACTACAGAGAGTCAAGATTTGATGATGATGATGATGATGATGAT 492
421 TAAAGAACTATGAGAGAACTACAGAGAGTCAAGATTTGATGATGATGATGATGATGATGAT 480
493 TTTCCCTTTTGTGAGTGTGAGCCAGTACTTAAATAACCTTTGTCTACAGCCCTCCG 552
481 TTTCCCTTTTGTGAGTGTGAGCCAGTACTTAAATAACCTTTGTCTACAGCCCTCCG 521
553 CTCTCTCTGCTGAGCAATTTGAAAAAGCATAGTGAAGACTTCTGTCTCTCTCTCTCTA 612
522 ----- 521
613 TGTGCTGTGTTATGTCAGAACTAAGTACCAAAATGATCTTATAGAGAGGTAATAAAA 672

Db 522 ----- 521
QY 673 TATGATCTATGCTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGAGTGGGA 732
Db 522 ----- 521
QY 733 TCAGTCTACAGTGAAGTTCTAGAGAGACCCACTACGTTATCTGAGCAATGGCAAAAAGC 792
Db 522 ----- 521
QY 793 TCAGATATGCTTATTTGAAAACCTACTGGGATTTTCAATTTCTTCAACCCACTTACCAA 852
Db 556 TGACATATGCTTATTTGAAAACCCCTGAGATTTTCAATTTCTTCAATTTCTTCAATTTCA 615
QY 853 TGTGAGTCTGTTGAGAGAGCTCCACTGCAACCTGCAACCCCTTACCCAGAGAAATGGA 912
Db 616 CGTTGATTTTGTGAGAGATTCACCTGCAACCTGCAACCCCTTACCCAGAGAAATGGA 675
QY 913 AGAGTTTGTCCAGAGCTCTGAGAGAAATGAGTGTGTTGTTTCTCTGGGGTGCATGAT 972
Db 676 GGAATTTGTACAGAGCTCTGAGAGAAATGAGTGTGTTTCTCTGGGGTGCATGAT 735
QY 973 CAGTACACGTCAGAGAGAGGCGCAATGTAATTCATCAGCCCTTGCAGATCCACA 1032
Db 736 AAGTACATGACAGAGAGAGGCGCAATGTAATTCATCAGCCCTTGCAGATCCACA 795
QY 1033 AAGGTTCTGTGAGATTTGATGGAAATTAACCATCTTTAGACTCAATCTAGCCCT 1092
Db 796 AAGGTTCTGTGAGATTTGACGGGAAATTAACCATCTTTAGACTCAATCTAGCCCT 855
QY 1093 GTACAGATGATATACCCAGAAATGATCTTCTGTCATCCCAAAACCAAGCTTTTATAC 1152
Db 856 GTACAGATGATATACCCAGAAATGATCTTCTGTCATCCCAAAACCAAGCTTTTATAC 915
QY 1153 TCATGATGAAATGAATGGGATCTATGAAGCTATTTTACATGGGGTCTTATGTTGGAGT 1212
Db 916 TCATGATGAAATGAATGGGATCTATGAAGCTATTTTACATGGGGTCTTATGTTGGAGT 975
QY 1213 TCCCATTTTGGTATCAGCTTGAATACATAGCTCAATGAAAGGCAAGAGAGAGCTGT 1272
Db 976 TCCCATTTTGGTATCAGCTTGAATACATAGCTCAATGAAAGGCAAGAGAGAGCTGT 1035
QY 1273 AGAAATTAACCTCAAACTATGACAGAGCAAGATTTTACAGAGGCTTTGAGAGAGCTAT 1332
Db 1036 TAGATTTGATCTTCAACATATGCAATGACAGAGCTGCTGATGACAGAGAGAGCTAT 1095
QY 1333 TACCGATTCCTCTTATTAAGAGAAATGCTATGAGATTTCAAGATTTCAACATGATCAAC 1392
Db 1096 TATGATCTCTTATTAAGAGAAATGCTATGAGATTTCAAGATTTCAACATGATCAAC 1155
QY 1393 TGTAAAGCCCTGATTCAGAGAGCTTCTGATTCAGAGTTTGTCAAGCCCAAGAGAGC 1452
Db 1156 AGTAAAGCCCTGATTCAGAGAGCTTCTGATTCAGAGTTTGTCAAGCCCAAGAGAGC 1215
QY 1453 CAAGACCTGAGATGAGCTGAGCCATGACCTCACTGTTCCAGAGCTTCACTATGATAT 1512
Db 1216 CAAGACCTGAGATGAGCTGAGCCATGACCTCACTGTTCCAGAGCTTCACTATGATAT 1275
QY 1513 GATTTGGTCTCTGATGAGCTGATGAGCACTGCTATTTCTTCACAAAATGTTTTT 1572
Db 1276 GATTTGGTCTCTGATGAGCTGATGAGCACTGCTATTTCTTCACAAAATGTTTTT 1335
QY 1573 ATTTTCTGTCAAAATTTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
Db 1336 GTTTTGTCTGAGAGTTTGTAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
QY 1633 AA 1634
Db 1396 GA 1397

RESULT 6
US-09-356-806-1

Sequence 1, Application us/09356806
Patent No. 6586175
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
TITLE OF INVENTION: 2B15 (UGT2B15) Genes
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
NUMBER OF FILING DATE: 1999-07-20
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1323
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: Other
LOCATION: (140)...(897)
US-09-356-806-1
Query Match 45.7%; Score 754.8; DB 4; Length 1323;
Best Local Similarity 99.7%; Pred. No. 9e-211;
Matches 756; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGCACTGGAAAAACAGATTTGATGCTATGAGATGCTATGAAATGAGACTTCACTCT 60
Db 143 AGCACTGGAAAAACAGATTTGATGCTATGAGATGCTATGAAATGAGACTTCACTCT 202
QY 61 TCTCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGAGTGTGGAAGAGCTGTGTG 120
Db 203 TCTCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGAGTGTGGAAGAGCTGTGTG 262
QY 121 GCCCAGAAATTCAGCAGCTGATGAATATTAAGACAATCTGATGAATCTTCCAGAG 180
Db 263 GCCCAGAAATTCAGCAGCTGATGAATATTAAGACAATCTGATGAATCTTCCAGAG 322
QY 181 AGCTATGAGGTGATCTGATTTGGCATCTTCAGCTTCCATTTCTTCCATCCCAAGGCC 240
Db 323 AGCTATGAGGTGATCTGATTTGGCATCTTCAGCTTCCATTTCTTCCATCCCAAGGCC 382
QY 241 ATCTACTCTTAATTTGAAGTTATCTGATCTTAACTTAACTGAGTTGAGATAT 300
Db 383 ATCTACTCTTAATTTGAAGTTATCTGATCTTAACTTAACTGAGTTGAGATAT 442
QY 301 TATCAGAGCTGTTAAGAGATGGCAGAACTTCAAAAGACATTTTGGTCAATTT 360
Db 443 TATCAGAGCTGTTAAGAGATGGCAGAACTTCAAAAGACATTTTGGTCAATTT 502
QY 361 TTCACAGTACAGAAATCATGATGACATTTAAAGACATCTTAAAGAGTCTGTAAGA 420
Db 503 TTCACAGTACAGAAATCATGATGACATTTAAAGACATCTTAAAGAGTCTGTAAGA 562
QY 421 TATGTTTCAATTAAGAACTATGAAGAACTACAGAGTCAAGATTTGATGTTGTCT 480
Db 563 TATGTTTCAATTAAGAACTATGAAGAACTACAGAGTCAAGATTTGATGTTGTCT 622
QY 481 TGCAGATGCTTTTCCCTTTGGTGAAGCTGCGCAGATTAATTAATACCTTTGT 540
Db 623 TGCAGATGCTTTTCCCTTTGGTGAAGCTGCGCAGATTAATTAATACCTTTGT 682
QY 541 CTACAGCTCGCTTCTCTCTGCTACGCAATTAAGAAAGCAATGATGAGAGACTTCTGT 600
Db 683 CTACAGCTCGCTTCTCTCTGCTACGCAATTAAGAAAGCAATGATGAGAGACTTCTGT 742
QY 601 CCTCTCTCTATGAGCTGTTGTTATGCTGAGACTAAGTCAACCAATGAGCTTATAGA 660
Db 743 CCTCTCTCTATGAGCTGTTGTTATGCTGAGACTAAGTCAACCAATGAGCTTATAGA 802
QY 661 GAGGTAATAATAATGATCTATGCTTATTTGAAATTTGGTCCAAATATTTGACAT 720

Db 803 GAGGTAATAATGATCTATGCTTATTTGAATTTGTTCCAAATATTGACAT 862
Qy 721 GAAGAGTGGATAGATTCTACATGATGAGTCTTGAA 758
Db 863 GAAGAGTGGATAGATTCTACATGATGAGTCTTGAA 900

RESULT 7

US-09-356-806-41
; Sequence 41, Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; FILE REFERENCE: SFO-22PRV2
; CURRENT APPLICATION NUMBER: US/09/356,806
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 1686
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (392)...(1126)
US-09-356-806-41

Query Match 37.0%; Score 610.8; DB 4; Length 1686;
Best Local Similarity 87.9%; Pred. No. 1.3e-168;

Matches 666; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 1 AGCACTCGAAGAAACAGATTGCTATGATAGATGCTATGAAATGAGACTTACGCT 60
Db 372 AGCACTCGAAGAAACAGATTGCTATGATAGATGCTATGAAATGAGACTTACGCT 431
Qy 61 TCTCCTGATGAGTGAAGTCTTACTTAACTGAGGAGTGGAAAGGCTGCTGTG 120
Db 432 TTGCTTAATCACTGAGCTTTTCTTAACTGAGGAGTGGAAAGGCTGCTGTG 491
Qy 121 GCCCAGCAATTCAGCCATGATGAAATATTAAGACATCTGATGAACTTGTCCAG 180
Db 492 GGCAGCAGAAATACAGCCATGATGAAATATTAAGACATCTGATGAGTATTCAG 551
Qy 181 AGGTATGAGGTGACTGATTTGGCATCTTCACTTCTTCCATCCCAAGCC 240
Db 552 AGGTATGAGGTGACTGATTTGGCATCTTCACTTCTTCCATCCCAAGCC 611
Qy 241 ATCTACTCTTAATTTGAAGTTATCTGATCTTAACTTAACTGAGTGGATGAT 300
Db 612 ATCCGCTCTTAATTTGAAGTTATCTGATCTTAACTTAACTGAGTGGATGAT 671
Qy 301 TATCAAGCAGGTGTTAAGATGGGAGAACTTCAAAAGACATTTGGTCAATTT 360
Db 672 CATCATGCAACAGTTAAGATGGTCAAGCTTCAAAAGATGATTTGGTCAATTT 731
Qy 361 TTCAAGATGACAGAAATCATGTGACATTTAATGACATCTTGAAGATTTCTTA 420
Db 732 TTCAAGATGACAGAAATCATGTGACATTTAATGACATCTTGAAGATTTCTTA 791
Qy 421 TATAGTTCAAAATAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 480
Db 792 TGTAGTTCAAAATAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 851
Qy 481 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGGCAGATTAATTAATCCCTTGT 540
Db 852 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGGCAGATTAATTAATCCCTTGT 911

Qy 541 CTACAGCTCCGCTCTCTCTGAGCTACGCAATTTGAAGACATAGTGAGAGACTTGT 600
Db 912 GTACAGTCTAGCTTCTCTCTGAGCTACGCAATTTGAAGACATAGTGAGAGACTTGT 971
Qy 601 CCTCTCTCTATGAGCTGCTGTTGATGACAGAACTAAGTACCAATGACTTCAAGA 660
Db 972 CCTCTCTCTAGTACCTGCTGTTGATGACAGAACTAAGTACCAATGACTTCAAGA 1031
Qy 661 GAGGTAAATAATGATCTATGCTTATTTGAAATTTGGTCCAAATATTGACAT 720
Db 1032 GAGGTAAATAATGATCTATGCTTATTTGAAATTTGGTCCAAATATTGACAT 1091
Qy 721 GAAGAGTGGATAGATTCTACATGATGAGTCTTGAA 758
Db 1092 GAAGAGTGGATAGATTCTACATGATGAGTCTTGAA 1129

RESULT 8

US-09-671-317-403
; Sequence 403, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marla
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguerelet, Lydie
; APPLICANT: Cohen, Amick
; TITLE OF INVENTION: BIALLERIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 403

LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 503
OTHER INFORMATION: 10-448-266 : polymorphic base A or C
NAME/KEY: misc_binding
LOCATION: 483..502
OTHER INFORMATION: 10-448-266.miel, potential
NAME/KEY: misc_binding
LOCATION: 504..523
OTHER INFORMATION: 10-448-266.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 238..257
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 660..679
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 491..515
OTHER INFORMATION: 10-448-266 potential probe
NAME/KEY: misc_feature
LOCATION: 772..907,915
OTHER INFORMATION: n=a, g, c or t
US-09-671-317-403

Query Match 34.7%; Score 572.4; DB 4; Length 1001;
Best Local Similarity 86.4%; Pred. No. 1.7e-157;
Matches 643; Conservative 1; Mismatches 97; Indels 3; Gaps 1;

FILE REFERENCE: 62.US3.CIP
CURRENT APPLICATION NUMBER: US/09/671.317
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/536,178
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/IB00/00403
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: US 60/126,269
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/131,961
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 977
SOFTWARE: Patent.pm
SEQ ID NO 412
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 503
OTHER INFORMATION: 10-457-284 : polymorphic base G or T
NAME/KEY: misc binding
LOCATION: 483..502
OTHER INFORMATION: 10-457-284.misl, potential
NAME/KEY: misc binding
LOCATION: 504..523
OTHER INFORMATION: 10-457-284.mis2, potential complement
NAME/KEY: primer bind
LOCATION: 220..238
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer bind
LOCATION: 621..639
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc binding
LOCATION: 491..515
OTHER INFORMATION: 10-457-284 potential probe
NAME/KEY: misc feature
LOCATION: 715
OTHER INFORMATION: n=a, g, c or t
US-09-671-317-412

Query Match 30.4%; Score 502.2; DB 4; Length 1001;
Best Local Similarity 80.3%; Pred. No. 5.8e-137;
Matches 612; Conservative 1; Mismatches 145; Indels 4; Gaps 2;
QY 1 AGCACTGGAAACACACATTCGATTCATCGAGATGCTTATGAATGACCTTCT 60
DB 217 AACAACTGAAAGAAAGACATTCGATTAAGACAGAGATGCTTGAATGACGTCAGCTT 276
QY 61 TCTCTGATCAGCTGAGCTGTTACTTACTGCTGGAGTTGGGAAGGCTGGTG 120
DB 277 TCTGCTGATCAGCTGAGCTGTTACTTACTGCTGGAGCTGGGAAGGCTGGTG 336
QY 121 GCCCAGCAATTCAGCCATCGATGATTAATGAAGCAATCTGGATGACCTTGCAG 180
DB 337 GCCCAGCAATTCAGCCATCGATGATTAATGAAGCAATCTGGATGACCTTGCAG 336
QY 181 AGCTCATGAGTGAATGATTTGGCATTTCACTTCTTTGATCCCAAGCCC 240
DB 397 GGGTCAATGAGTGAATGATTTGGCATTTCACTTCTTTGATCCCAAGCCC 456
QY 241 ATCTACTTAATTTGAAATTTATCTGATCTTTTAACTTAATGAGTTGAGATAT 300
DB 457 ATCTACTTAATTTGAAATTTATCTGATCTTTTAACTTAATGAGTTGAGATAT 516
QY 301 TATCAAGCAGCTGGTTAAGATGGGCA--GAACCTCAAAAGACATTTTGTCATA 357
DB 517 TCTTCAAAATCTGATGATGATGATATGATGTTTCAAAATATCATTTTGTGATA 576
QY 358 TTTTTCACAGTACAGAAATCATGTGACATTTAATGACATCTTGAAGTTCTGTAA 417
DB 577 TTTTTCACATTTCAAGAAATTTGTTGGGAATTTATGACTACAGTAAAGCTCTGTAA 636

QY 418 GATATAGTTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATTTGT 477
DB 637 AGATGACGTTTGAATAGAACTTATGATGAAGAACTACAGAGTCAAGATTTGATTTGT 696
QY 478 TCTTGCAGATGCTGTT-TTCCCTTTGGTGAAGTCTGCGCGAGTTACTTAATATCCCT 536
DB 697 TCTGCGAGATGCCCTTAATNTCCCTGGTGTAGCTACGCTGGAATTAATATATCCCT 756
QY 537 TTGTCTACAGCTCCGCTTCTCTGCTGCTAGCAATTAAGAACTGAGGAGGACTTC 596
DB 757 TTCTGACAGCTCTTCAATTTCTGTTGCTACACATTTGGAAGAAATGGGAGGATTC 816
QY 597 TGTTCCTCTCTTCTTATGCTGCTGTTTATGTCAGACTAAGTCAATGACTTCA 656
DB 817 TGTTCCTCTCTTCTTATGCTGCTGTTTATGTCAGACTAAGTCAATGACTTCA 876
QY 657 TGAAGAGGTAAATAATGATCTATGCTTTATTTGATTTGTTCCAAATATTTG 716
DB 877 TGAAGAGGTAAATAATGATCTATGCTTTATTTGATTTGTTCCAAATATTTG 936
QY 717 ACATGAAGAGTGGATCAGTTCTACAGTCAAGTCTTAGAA 758
DB 937 ATCTGAAGAGTGGACAGTTTATGTAAGTCTTAGGA 978

RESULT 11

Sequence 2, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2339 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US92-00282-2

Query Match 19.8%; Score 327.4; DB 5; Length 2339;
Best Local Similarity 52.9%; Pred. No. 1.2e-85;
Matches 806; Conservative 0; Mismatches 701; Indels 18; Gaps 4;
QY 65 CTGATACAGTACGCTGTTACTTATGCTGGAGTTGGAAGGCTGCTGGGCC 124

46 CTGCTGCTCCTCCTCAGTGTCCAGCCCTGGGCTGAGTGAAGAAAGTGTGTGTGTCCTCC 105
QY 125 ACAGAAATTCAGCCATGATGATATTAAGACAATCTGTGATGAACTTTGCCAGAGGT 184
Db 106 ACTGATGAGGAGCCCTGCTCAGCATGCGGAGGCTTGGGGGCTTCATGCCAGAGGC 165
QY 185 CATGAGGTGACTGATTTGGGACTTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCT 244
Db 166 CACGAGGCGGTGCTTCACCCGAGAGTGAATTCACATCAAGAAAGAAATTTTTC 225
QY 245 ACTCTTAATTTGAAGTTATCTGTATCTTTAACTAAATGAGTTGAGATATATC 304
Db 226 ACCCTGA-----CAGCTATGCTGTTCATGACCCAGAAAGAAATTTTCGCTTA-- 277
QY 305 AAGCAGCTGTTAAGATGGGCGAAGCTTCCAAAGACATTTTGTCTATTTTCA 364
Db 278 ----CGCTGGGCTCAGCTCAAGGGTCTTTGAAACAGAACTCTTGAAGAGATTTCT 333
QY 365 CAAGTACAAAGAAATCATGTGACATTTAATAGACATCTTAAGAAAGTTCTGTAAGATATA 424
Db 334 AGAAGTATGGCAATTAATGAACAATGTATCTTGGCCCTTCATAGTGTGTGTGAGCTA 393
QY 425 GTTTCAAATAGAAGAACTTATGAGAAAGTACAGAGTCAAGATTTGATTTTGTGCA 484
Db 394 CTGATATATGAGGCGCTGATCAGACCTGAATGCTACTTCTTGTATGTGTTTAA 453
QY 485 GATGCTGTTTCCCTTTGTGAGTGTGCTGCGGAGTTACTTAAATACCTTTGTCTAC 544
Db 454 GACCCGTTAAGCTGCGGGGCGGTGCTGCTAAGTCACTGCAATTCGCTGTGTT 513
QY 545 AGCTCGCTCTCTCCTGGGCTAGCAATGTAAGAAAGTATGAGAGACTTCTGTCCCT 604
Db 514 TTTGAGAGTACATTCATGTGAC--TTAGCTTTAAGGACACAGTGTCCAAATCT 570
QY 605 CTTCTCTATGTCCTGTGTATGTACAGAACTAAGTGAACAATGACTTTTATAGAGG 664
Db 571 TCTCTATATTTCTTAAGTACTAAGCAATTCAGACCATGACATTCCTGGAAGG 630
QY 665 GTAAAAATATGATCTATGTGCTTTATTTGAATTTTGTCCAAATTTTGAAGTGAAG 724
Db 631 GTCAAGAACTGCTCACTGCTGCGCTGCTCAATTTGCAATCTTTTCTGCCCCCT 690
QY 725 AAGTGGATCAAGTCTACAGTGAAGTTCTAGGAAAGCAACCACTAGTATCTGAGCAATG 784
Db 691 TATGCAAGCTC--TGCTCTGAGCTTTTTCAGAGAGAGTGTCACTGTGATCTTGTG 747
QY 785 GCAAAAGCTGACATATGCTTAATGCAAACTACTGGATTTTCAATTTCTCACCCTC 844
Db 748 AGCTATGATCCGTGTGCTGTTCGAGGGGACTTTGTGATGAGACTAACCCGAGCGCATC 807
QY 845 TTACCAAAATGTTGATGTTGGAGACTCCACTGCCAACTGCGCAACCCCTACCGAAG 904
Db 808 ATGCCCAATGCTCTTCAATGAGGGGCAATCACTGTGCCAAGGGGAGCACTATCTCAG 867
QY 905 GAAAGGAAGATTTGTCCAGAGCTGTGAGAAATGTTGTGTGTTTCTCTGGGG 964
Db 868 GAATTTGAAGCTCAATTAATCTCTTGAGAAACAGATGTGTTTCTCTTTGGGA 927
QY 965 TCGATGTCATGAACAAGTCAAGAGAAAGGCAATGTAATTCATCAGCCCTTGGCAAG 1024
Db 928 TCAATGTGTCAGAAATTTCAAGAAAGAAAGTATGAGCAATGCTGTGAGCAAT 987
QY 1025 ATCCCAAAAGGTTCTGTGAGATTTGATGGAAATTAACCAAGATCTTTAGACTCAAT 1084
Db 988 ATCCCTCAGACAGTCTGTGGGGGTCACTGGAACCCGACCAATGCAATCTTGGAAAC 1047
QY 1085 ACTGCGCTGTAAGAGTGAATCCCAAGATGATCTTCTGTGATCCCAAAAGGCT 1144
Db 1048 AGGATATCTGTTAAGTGGCTACCCCAAAAGATGCTGTGTCACCCGATGACCGGTGC 1107
QY 1145 TTTATCACTCATGTGGAATGAATGGAATCTATGAAGCTATTTACCATGGGTCTCTATG 1204
Db 1108 TTTATCACTCATGTGTCCTTCCATGTGTTTATGAAGATATGCAATGGCGTGTCCCATG 1167

QY 1205 GTGGAGTTCCCATATTTGGTGTATCAGCTTGATTAACATAGCTCATGAAGGCCAAAGGA 1264
Db 1168 GTATATAGTCCCTGTTTGGTGTATCATGTGAACAATGCAAAAGCGCATGAGACTTAAGGA 1227
QY 1265 GCACTGTGAATTAATACTTCAAAACTATGACAAAGGAAGATTTACTGAGGGCTTTGAGA 1324
Db 1228 GCTGAGATGACCTGTAATGTCTGGAATGACTTCTGAATTTAGAAAATGCTTAA 1287
QY 1325 ACAGTATTCAGATTCCTTATTAAGAGATGTATGATGATTAACAAGATTAACCAT 1384
Db 1288 GCACTCATATGACAAAGTTTCAAGAGAAACATCAGGGCTCTCCAGCTTCAAG 1347
QY 1385 GATCAACTGTAAAGCCCTAGATCAGAGAGCTTTGATGAGTTGTATGCGCCAC 1444
Db 1348 GACCGCGGTGAGCGCGCTGAGACTGCGCGTGTCTGGGTGAGTTGTATGAGGCGAC 1407
QY 1445 AAGAGCCAGACCTGCGATCAGCTGCCATGACTCACTGCTGTTCCAGCACTACTCT 1504
Db 1408 AAGGCGCGCAGACCTGCGCGCCCGCAGCCCAAGCCTCACTGTATCCAGTACATCTC 1467
QY 1505 ATGATGTGATGGGTTCTGCTGACCTGTGTGCAACTGCTATATCTGTTCACAAA 1564
Db 1468 TTGAGCTGATGTGTTCTCTTGGCCGTGTGCTGACAGTGGCTTCACTTAA 1527
QY 1565 TGTTTTATTTTCTGTCAAAAT 1589
Db 1528 TGTGTCTATGTGCTACCGGAAT 1552

RESULT 12
PCT-US92-00282-1
Sequence 1, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGI1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2336 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US92-00282-1
Query Match 19.2%; Score 316; DB 5; Length 2336;

Best Local Similarity 55.2%; Pred. No. 2.5e-82;
Matches 660; Conservative 0; Mismatches 530; Indels 6; Gaps 2;

```

QY 394 TGACATCTAGAAAAGTTCTGTAGATATAGTTTCAAAATAGAACTTAAAGAACT 453
Db 360 TGCTATGCTTTTGTCTGCTGCTTCCCACTTACACAAAGAGCTCAGGCTCCCT 419
QY 454 ACAGAGTCAAGATTTGATGTTCTTGTGAGATGCTGTTTCCCTTGGTGAAGTCT 513
Db 420 GGCAGAAAGAGCTTTATGTCATGCTGAGGAGCCCTTCTCTTGTGAGCCCACT 479
QY 514 GGCAGATTAATAAATACCCTTGTCTACAGCTCCGCTTCTCTGCTGAGCAAT 573
Db 480 GGCCTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536
QY 574 TGAAGACATAGTGGAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 633
Db 537 GGAATTTGAGAGCTTCCAGTCCCAAGCCCAATTTCTCTGCTGCTGCTGCTGCT 596
QY 634 ACTAAGTACCAATGATCTTATAGAGAGGTAATAATATGATCTATGCTTATTT 693
Db 597 TCATTCATGATCAGTACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 656
QY 694 TGAATTTGCTTCCAAATATTTGACATGAAAGTGGATTCAGTCTACAGTGAAT 753
Db 657 GAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 713
QY 754 AGGAAGCCCACTGCTATCTGAGACATGCAAAAGTGAATGCTATGCTATGCA 813
Db 714 TCAGAGAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 773
QY 814 CTACTGGATTTTCAATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 873
Db 774 TGACTTTGTAAGATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
QY 874 CCACTGCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 933
Db 834 CAATGCTTCAACCAAAATCACTATCCAGAAATTTGAAGCTTCACTATGCTT 893
QY 934 AGAAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993
Db 894 AGAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 953
QY 994 GGCATATGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1053
Db 954 AGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1013
QY 1054 TGGGAATTAACAGATATCTTGAAGTCAATCTGCTGCTGCTGCTGCTGCTG 1113
Db 1014 TGGAACTGCAATCGAATCTTGGGAAACAGATCTGTTAAAGTGGCTACCC 1073
QY 1114 TGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1173
Db 1074 CGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
QY 1174 CTATGAAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1233
Db 1134 TTATGAAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1193
QY 1234 TGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1293
Db 1194 GGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1253
QY 1294 GACAAAGAAATTTACTGAGGCTTTGAGAACATGATTAACGATCTCTTT 1353
Db 1254 GACTTCTGAAATTTTAAAGATGCTTAAAGAGATGATCAATGCAAAAGTT 1313
QY 1354 GAATGCTATGATTAATCAAGAAATTCATGATCAACCTGTAAGGCTTGAAG 1413
Db 1314 GAACATAGAGGCTCTTCAAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 1373
QY 1414 AGCTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1473

```

```

Db 1374 CGTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1433
QY 1474 CCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1533
Db 1434 CCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1493
QY 1534 TGTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1589
Db 1494 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1549

```

RESULT 13

```

US-09-305-856B-17
; Sequence 17, Application US/09305856B
; Patent No. 6479236
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; TITLE OF INVENTION: Genotyping the Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 1 (UGT1) Gene
; FILE REFERENCE: 4389-7 (formerly SEQ-17C1P)
; CURRENT APPLICATION NUMBER: US/09/305,856B
; CURRENT FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 60/084,807
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(735)
US-09-305-856B-17

```

Query Match 16.3%; Score 269; DB 4; Length 735;

Best Local Similarity 62.0%; Pred. No. 7.9e-69;

Matches 425; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

```

QY 905 GAAATGAAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 964
Db 1 GAAATGAAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 965 TCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
Db 61 TCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 1025 ATCCCAAAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1084
Db 121 ATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 1085 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144
Db 181 AGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 1145 TTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
Db 241 TTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 1205 GTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1264
Db 301 GTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 1265 GCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1324
Db 361 GCTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 1325 ACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
Db 421 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 1385 GATCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1444

```


Query Match	14.7%;	Score 242.8;	DB 4;	Length 1001;
Best Local Similarity	76.8%;	Pred. No. 4.2e-61;		
Matches 322;	Conservative 0;	Mismatches 93;	Indels 4;	Gaps 2;

```

Oy      1  AGCAACTGGAAAAACAAGCAATTGCATTGCATCAGAGATGTCATATGAATAGGACT--TCAGCTC 5/9
Db      583  AGCCATTGGAAAAAGACATTGCATTGCACACAGAGAACTATATATAATCGATMGCAATA 64/22
Oy      60  TTCTCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGGAAAAAGTCTGTGT 11/9
Db      643  TTCTCTGATACACTCACTTGTTCATTTTAGTTCTGGAGTTGTGGAAAGTCTGTGT 70/22
Oy      120  GGCCCAACAGAAATTCAGCCACTGATGATATTTAAAGACATCTGGATGAACTTGTCCAGA 17/9
Db      703  GGCCCAACAGAAATTCAGCCACTGATGATATTTAAAGACATCTGGATGAACTTGTGGGA 76/22
Oy      180  GAGCTCATGAGGAGCATGTATTGGCATCTCAGCTTCATCTTCGATCCCAACAGCC 23/9
Db      763  GAGGCGATTAGATGACTGTGCTGTCAATCTTTGGCTTCATTATTATTCACCTTAGTAAT 82/22
Oy      240  CATCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTMAAACCTGAGTTGAGGATA 29/9
Db      823  CATCTGCTATTGATTTGAGGTTATCCAAACATCTCTAATTTAAATTAATTTGAAGGTC 88/22
Oy      300  TTATCAAGACGCTGGTTAAGATGGGCA---GACTTCCAAAAGACACATTTTGGTCAT 35/6
Db      883  TTGTGTGTAAGTGAATCAATAGATGATATACATCTTCGAAAAGATGCATTTTGGTCAT 94/22
Oy      357  ATTTTTCACAAAGTACAGAAATCATGTGGACATTTATGACATACTTAGAAAGTTCGT 41/5
Db      943  ATTTCTCACAGACACAGAACTCTTTTGGAACTCTACAGACTGTGTTAATTAATCTCTGT 100/12

```

Search completed: December 7, 2003, 07:39:50
Job time : 95 secs

THIS PAGE BLANK (USPTO)

|||||
Db 241 CGCTCTAAAATTTAAATTTATCCACATCTTTAACTAAAATGAGTTGGAGAAATTCAT 300
Qy 304 CAACAGCTGGTTAAGATGGGAGAACTTCAAAAAGACATTTGGTCATATTTTTC 363
Db 301 CATGCAACAGATTAAAGATGGTCAAGACTTCCAAAAGATACATTTGGTATATTTTTC 360
Qy 364 ACAAGTACAGAAATTCATGTGAGATTTATGACATCTTAGAAAGTCTTAGAGATAT 423
Db 361 ACAAGTACAGAAATTCATGTGAGATTTATGAGATTTAGAAAGTCTTAGAGATAT 420
Qy 424 AGTTCAATTAAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTGTTTC 483
Db 421 AGTTCAATTAAGAACTTATGAAAGAAAGTCAAGAGTCAAGATTTGACGTCATTTTTC 480
Qy 484 AGATGCTGTTTCCCTTGGTGGAGCTGGCCGAGTACTTAAATACCTTTGCTA 543
Db 481 AGATGCTATTTTCCCTGAGTGGAGCTGGCTGAGCTATTTAACTAACCTTTGATGA 540
Qy 544 CAGCTCCGCTTCTCTCGTGGCTACGCAATGAAAGATAGTGGAGACTTCTGTTCCC 603
Db 541 CAGTCTCAGCTTCTCTCGTGGCTACATTTGAAAGATAGTGGAGATTTATTTTCCC 600
Qy 604 TCCTTCTTANGTGCCTGTTGTATGTCAGAACTAAGTACCAATGACTTTTCATAGAG 663
Db 601 TCCTTCTTACGTACCTGTTGTATGTCAGAACTAAGTACCAATGACTTTTCATAGAG 660
Qy 664 GGTAAAAATATGATCTATGCTGTTATTTTGAATTTTGGTCCAAATTTTGACATGAA 723
Db 661 GGTAAAAATATGATCTATGCTGTTATTTGACTTTTGGTCCAAATTTTGACATGAA 720
Qy 724 GAATGGAGTCAATCTACAGTGAAGTCTTAGAAGACCCACTAGTTATCTGAGACAT 783
Db 721 GAATGGAGTCAATCTATGATGAAGTCTTAGAAGACCCACTAGTTATCTGAGACAT 780
Qy 784 GCGAAAAGTGAATATGCTTATTTGAACTAGTGGATTTTCAATTTCTCAACCACT 843
Db 781 GCGAAAAGTGAATATGCTTATTTGAACTAGTGGATTTTCAATTTCTCAACCACT 840
Qy 844 CTATACCAATATGATGCTGTTGAGAGTCACTGCAAACTGCAAACTGCAAACTGCAAA 903
Db 841 CTATACCAATATGATGCTGTTGAGAGTCACTGCAAACTGCAAACTGCAAACTGCAAA 900
Qy 904 GGAATGGAAGAGTTTGTCCAGAGCTCTGAGAGAAATGATGTTGGTGTCTCTGGG 963
Db 901 GGAATGGAAGAGTCTTGTACAGAGCTCTGAGAGAAATGATGTTGGTGTCTCTGGG 960
Qy 964 GTCGATGTCAGTAAACGTCAGAGAAAGGCAATGTAATTCATCAGCCCTTGGCAA 1023
Db 961 GTCGATGTCAGTAAACGTCAGAGAAAGGCAATGTAATTCATCAGCCCTTGGCAA 1020
Qy 1024 GATCCCAAAAAGTTCTGTGAGAGTGTGATGGAATTAACCAAGATCTTTAGAGATCTAA 1083
Db 1021 GATCCCAAAAAGTTCTGTGAGAGTGTGATGGAATTAACCAAGATCTTTAGAGATCTAA 1080
Qy 1084 TACTCGGCTGACAGAGTATCCCAAGATGATCTTGGTGTATCCCAAAACCAAGC 1143
Db 1081 TACTCGGCTGACAGAGTATCCCAAGATGATCTTGGTGTATCCCAAAACCAAGC 1140
Qy 1144 TTTTATCACTCATGATGGAATGAAATGGAATCTATGAACTATTTTACATGGGCTCTAT 1203
Db 1141 TTTTATCACTCATGATGGAATGGAATGGAATCTATGAACTATTTTACATGGGCTCTAT 1200
Qy 1204 GGTGGAGTTTCCATATTTGGTGTATCAGCTTGTATTAACATAGACTCAATGAAGGCCAAAG 1263
Db 1201 GGTGGAGTTTCCATATTTGGTGTATTAACATAGACTCAATGAAGGCCCAAGGG 1260
Qy 1264 AGCAGCTGTAGAAATTAACCTCAAACTAGTGAAGAGATTACTGAGGGCTTTGAG 1323
Db 1261 AGCAGCTGTAGAGTGAATTAACCAATGTGAGTGAAGAGCTTGTGAAATGCAATTTGAA 1320
Qy 1324 AACAGTCAATTAACCAATCTCTTATTAAGAGATGCTATGAGATTAATCAAGATTTCA 1383
Db 1321 AACAGTCAATTAACCAATCTCTTATTAAGAGATGCTATGAGATTAATCAAGATTTCA 1380

Db 1321 GAGAGTAAATTAATGATCTCTCATATAAAGAAATGTTATGAAATTAATCAAGATTTCA 1380
Qy 1384 TGATCAACCTGTAAAGCCCTAGATGAGAGTCTTGTGATGAGTTGTGATGAGCCA 1443
Db 1381 TGATCAACCTGTAAAGCCCTAGATGAGAGTCTTGTGATGAGTTGTGATGAGCCA 1440
Qy 1444 CAAAGAGCCAGACCTCTGATGAGTCCATGAGTCCATGACCTGATGAGTCCATGAGC 1503
Db 1441 CAAAGAGCCAGACCTCTGATGAGTCCATGAGTCCATGACCTGATGAGTCCATGAGC 1500
Qy 1504 TATAGATGATGAGTGGTCTGCTGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1563
Db 1501 TTTGATGATGATGAGTGGTCTGCTGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1560
Qy 1564 ATGTTTATTTATTTTCCGTCGCAAAAATTAATTAAGTGAAGTGAAGAGGAGATA 1623
Db 1561 ATGTTTATTTATTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1620
Qy 1624 GATCTTTC 1631
Db 1621 GTTATATC 1628

RESULT 5
US-09-880-107-2120
; Sequence 2120, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2120
; LENGTH: 1855
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J05428
; US-09-880-107-2120

Query Match 75.9%; Score 1252.6; DB 10; Length 1855;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 21 TGCAATGATCAGAGATGTCATGAAATGACCTTCTCTCGATACAGTGAAGT 80
Db 1 TGCAATGATCAGAGATGTCATGAAATGACCTTCTCTCGATACAGTGAAGT 60
Qy 81 GTTACTTACCTTGGAGAGTTGGAAGAGTGTGTTGGGCCCAAGAAATTCAGCCACT 140
Db 61 TTTGCTTACCTCTGGGAATTTGGAAGAGTGTGTTGGGCCCAAGAAATTCAGCCACT 120
Qy 141 GATGAAATTAAGACAAATCTGGAATGAATTTGTCAGAGAGTCAATGAGTGAATCTGTAT 200
Db 121 GATGAAATTAAGACAAATCTGGAATGAATTTGTCAGAGAGTCAATGAGTGAATCTGTAT 180
Qy 201 TGCAATCTCAGCTTCATTTCTTTGATCCCAAGCCCATCTACTTAAATTTGAAG 260
Db 181 TGCAATCTCAGCTTCATTTCTTTGATCCCAAGCCCATCTACTTAAATTTGAAG 240
Qy 261 TTTATCTGATCTTTAATTAAGTGAATTTGAGATTAATTAAGAGAGTGTGAAG 320
Db 241 TTTATCCCACTCTTAACTAAAGTGAATTTGAGATTAATTAAGAGAGTGTGAAG 300

QY 321 GATGGGAGACTCCAAAAGACATTTTGTGATATTTTCACAAGTACAGAAATCA 380
 Db 301 GATGGTACAGCTTTCCAAAAGATACATTTTGTGATATTTTCACAAGTACAGAAATCA 360
 QY 381 TGTGACATTTTAATGACATCTAGTAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 440
 Db 361 TGTCAATTTTGTGACATCTAGTAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 420
 QY 441 TTATGAGAAATCTACAGAGTCAAGATTTGTGTTCTTTCAGATGCTGTTTCCCT 500
 Db 421 TTATGAGAAATCTACAGAGTCAAGATTTGTGTTCTTTCAGATGCTGTTTCCCT 480
 QY 501 TTGGGAGCTGCTGGCGAGTTACTTAATATACCTTTGTCTACAGCTCCGCTCTCTC 560
 Db 481 GTAGGAGCTGCTGGCGAGTTACTTAATATACCTTTGTCTACAGCTCTCTCTCTC 540
 QY 561 CTGGCTACGCAATGAAAGCATAGTGGAGGACTTCTCTCTCTCTCTCTCTCTCTG 620
 Db 541 CTGGCTACGCAATGAAAGCATAGTGGAGGACTTCTCTCTCTCTCTCTCTCTCTG 600
 QY 621 TTGTATGTCAAGTAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 680
 Db 601 TTGTATGTCAAGTAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 660
 QY 681 ATGTGCTTATTTTGAATTTTGTGTTCCAAATATTTGACATGAGAGTGGAGTCTCT 740
 Db 661 ATGTGCTTATTTTGAATTTTGTGTTCCAAATATTTGACATGAGAGTGGAGTCTCT 720
 QY 741 ACAGTGAATTTTGAAG 800
 Db 721 ATAGTGAATTTTGAAG 780
 QY 801 GAGCTTATTTTGAAG 860
 Db 781 GAGCTTATTTTGAAG 840
 QY 861 TGTGTTGAG 920
 Db 841 TGTGTTGAG 900
 QY 921 TCCAGAGCTCTGAG 980
 Db 901 TACAGAGCTCTGAG 960
 QY 981 GGTGAG 1040
 Db 961 TGACAG 1020
 QY 1041 TGTGAGAGATTTTGAAG 1100
 Db 1021 TGTGAGAGATTTTGAAG 1080
 QY 1101 GGATACCCAG 1160
 Db 1081 GGATACCCAG 1140
 QY 1161 GAATGAATGAGATCTATGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1220
 Db 1141 GAGCCAAATGAGATCTATGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1221 TTGGATGATGAGATCTATGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1280
 Db 1201 TTGGATGATGAGATCTATGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1281 ACTTCAAAATCTATGAGAGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1340
 Db 1261 ACTTCAAAATCTATGAGAGAGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1341 CCTCTTATTAAGAGAGATGATGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1400
 Db 1321 CCTCTTATTAAGAGAGATGATGAGATTTTACATGAGAGAGAGAGAGAGAGAGAGAG 1380

QY 1401 CCTAGATGAGAGAGATCTCTGATGAGATTTTGTATGAGAGAGAGAGAGAGAGAGAG 1460
 Db 1381 CCTAGATGAGAGAGATCTCTGATGAGATTTTGTATGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1461 TGGGATGAG 1520
 Db 1441 TGGGATGAG 1500
 QY 1521 TCCGCTGAG 1580
 Db 1501 TCCGCTGAG 1560
 QY 1581 GTCAAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 1631
 Db 1561 TGTGAAATTTTGTCTAG 1611

RESULT 6
 US-10-205-522-39
 ; Sequence 39, Application US/10205522
 ; Publication No. US20030077629A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penny, Laura
 ; APPLICANT: Galvin, Margaret
 ; APPLICANT: Miller, Andrew
 ; APPLICANT: Reidy, Michael
 ; TITLE OF INVENTION: Genotyping Human
 ; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
 ; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
 ; FILE REFERENCE: SEQ-22PRV2
 ; CURRENT APPLICATION NUMBER: US/10/205,522
 ; CURRENT FILING DATE: 2002-07-24
 ; PRIOR APPLICATION NUMBER: US/09/356,806
 ; PRIOR FILING DATE: 1999-07-20
 ; NUMBER OF SEQ ID NOS: 164
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 39
 ; LENGTH: 1854
 ; TYPE: DNA
 ; ORGANISM: H. sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (15) ... (1584)
 ; US-10-205-522-39

Query Match 75.6%; Score 1247.8; DB 14; Length 1854;
 Best Local Similarity 85.9%; Pred. No. 0;
 Matches 1384; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 21 TGCATGATGAGAGATGCTATGAAATGAGATTTGAGATCTCTCTGATAGAGTGAAGT 80
 Db 1 TGCATGATGAGAGATGCTATGAAATGAGATTTGAGATCTCTCTGATAGAGTGAAGT 60
 QY 81 GTTACTTAACTCTGGAGAGTGTGGAAGTGTGTTGTGTTGAGAGAGAGAGAGAGAG 140
 Db 61 TTTGCTTAACTCTGGAGAGTGTGGAAGTGTGTTGTGTTGAGAGAGAGAGAGAGAG 120
 QY 141 GAGTGAATTAAGAGAGATCTGATGAATCTTGTCCAGAGAGAGAGAGAGAGAGAGAG 200
 Db 121 GAGTGAATTAAGAGAGATCTGATGAATCTTGTCCAGAGAGAGAGAGAGAGAGAGAG 180
 QY 201 TGGCATCTTGAAGCTTCCATTTCTTTCGATCCCAAGCCCACTACTTAAATTTGAAG 260
 Db 181 TGGCATCTTGAAGCTTCCATTTCTTTCGATCCCAAGCCCACTACTTAAATTTGAAG 240
 QY 261 TTTATCTGTAATCTTAACTAAACTGAGTTGAGAGATTTATCAAGAGAGAGAGAGAG 320
 Db 241 TTTATCTGTAATCTTAACTAAACTGAGTTGAGAGATTTATCAAGAGAGAGAGAGAG 300
 QY 321 GATGGGAG 380
 Db 301 GATGGGAG 360

QY 381 TGTGACATTAAATGACACTAGTAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 440
DB 361 TGTCAATATTTGGGACATACCTAGTAAAGTCTGTAAAGATATAGTTTCAATTAAGAAC 420
QY 441 TTATGAAGAACTACAGAGTCAAGATTTGATTTGTTCTTGCAAGTCTGTTTCCCT 500
DB 421 TTATGAAGAAAGTCAAGAGTCAAGATTTGATTTGTTCTTGCAAGTCTGTTTCCCT 480
QY 501 TTGGTACGCTGCGGAGTCTTAAATACCTTTGTCTAGAGCTCCGCTTCTC 560
DB 481 GTATGAGCTGCTGAGCTATTTAATACCTTTGTCTAGAGCTCCGCTTCTC 540
QY 561 CTGGCTACGCAATTAAGAGATAGTGAAGAGTCTTCTCTCTCTCTCTCTCTCTG 620
DB 541 CTGGCTACGCAATTAAGAGATAGTGAAGAGTCTTCTCTCTCTCTCTCTCTCTG 600
QY 621 TTGTTATGTCAGAACTAAGTCAATGACTTTCAATGAGAGGTTAAATATGATCT 680
DB 601 TTGTTATGTCAGAACTAAGTCAATGACTTTCAATGAGAGGTTAAATATGATCT 660
QY 681 ATGTCCTTATTTGATTTGTTCCAAATATTTGACATGAAGAGGATGATCT 740
DB 661 ATGTCCTTATTTGATTTGTTCCAAATATTTGACATGAAGAGGATGATCT 720
QY 741 ACAGTGAAGTCTAGAGAACCCACTAGTATCTGACAAATGCAAAAGCTGACATAT 800
DB 721 ATAGTGAAGTCTAGAGAACCCACTAGTATCTGACAAATGCAAAAGCTGACATAT 780
QY 801 GGGTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCACTTTACCAATGTTAGT 860
DB 781 GGGTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCACTTTACCAATGTTAGT 840
QY 861 TCGTTGAGAGCTGACCTGCAACCTGCAACCCCTTACCAAGAAATGGAAGTTG 920
DB 841 TTTGTTGAGAGCTGACCTGCAACCTGCAACCCCTTACCAAGAAATGGAAGTTG 900
QY 921 TCCAGAGCTCTGAGAAATGTTGTTGTTGTTTCTGAGGTCGATGTTGATGATCA 980
DB 901 TACGAGAGCTCTGAGAAATGTTGTTGTTGTTTCTGAGGTCGATGTTGATGATCA 960
QY 981 CGTCAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGTCCCAAAAGTTTC 1040
DB 961 TGAAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGTCCCAAAAGTTTC 1020
QY 1041 TGTGAGATTTGATGAGAAATTAACCAAGTATTTGATGATGATGATGATGATGAT 1100
DB 1021 TGTGAGATTTGATGAGAAATTAACCAAGTATTTGATGATGATGATGATGATGAT 1080
QY 1101 GGATACCCCAAGATGATCTTCTGATCTTCCCAAAAGCTTTTATCACTGATGATG 1160
DB 1081 GGATACCCCAAGATGATCTTCTGATCTTCCCAAAAGCTTTTATCACTGATGATG 1140
QY 1161 GAATGAATGGAGTATGATGATTTTACCAATGAGGTCCTTATGATGAGGATTTCCATAT 1220
DB 1141 GAGCAATGGAGTATGATGATTTTACCAATGAGGTCCTTATGATGAGGATTTCCATAT 1200
QY 1221 TTGTTGATCAGCTGATGATTAACCTGATCAATGAAGGCGCAAGGCTGTAAATTA 1280
DB 1201 TTGTTGATCAGCTGATGATTAACCTGATCAATGAAGGCGCGAGGCTGTAAATTA 1260
QY 1281 ACTTCAAACTATGACAGAGATTTTACGAGGCTTTGAGAAAGCTTCACTGATGAT 1340
DB 1261 ACTTCAAACTATGACAGAGATTTTACGAGGCTTTGAGAAAGCTTCACTGATGAT 1320
QY 1341 CTTCTTAAAGAGATGATTAAGATTTTACCAATGATCACTGATGATCACTGATGAT 1400
DB 1321 CTTCTTAAAGAGATGATTAAGATTTTACCAATGATCACTGATGATCACTGATGAT 1380
QY 1401 CCGTGAATGAGAGCTCTTCTGATGATGATTTGATGATGATGATGATGATGATGAT 1460
DB 1381 CCGTGAATGAGAGCTCTTCTGATGATGATTTGATGATGATGATGATGATGATGAT 1440
QY 1461 TGGCATGAGTGGCCATGATCTGATGATGATGATGATGATGATGATGATGATGAT 1520

DB 1441 TTGGGTTGAGGCCAGACCTGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1521 TCTGCTGACCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGAT 1580
DB 1501 TCTGCTGACCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1581 GTCAAAATTTAATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 1631
DB 1561 TCTGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1611

RESULT 7

US-10-198-846-13134
Sequence 13134, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steilmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MRI-049
CURRENT FILING DATE: 2002-07-18
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13134
LENGTH: 2844
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 2824, 2825, 2826, 2827, 2828, 2829, 2830, 2831, 2832, 2833,
LOCATION: 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843,
LOCATION: 2844
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-13134

Query Match: 75.0%; Score 1237.2; DB 14; Length 2844;
Best Local Similarity 85.5%; Pred. No. 0;
Matches 1377; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 22 GCATTGATGAGATGCTATGAAATGAGACTTCTCTCTGATAGAGTGAAGT 81
DB 21 GCATTGATGAGATGCTATGAAATGAGACTTCTCTCTGATAGAGTGAAGT 80
QY 82 TTAATTTAGCTGTGGAGTTGTGAAAGTGTGTGTGTGCGCAGAAATTCAGCACTG 141
DB 81 TTAATTTAGCTGTGGAGTTGTGAAAGTGTGTGTGTGCGCAGAAATTCAGCACTG 140
QY 142 GATGAATTAAGAAATCTGTGATGAACTTGTCCAGAGGTCATGAGTGTGATGAT 201
DB 141 GATGAATTAAGAAATCTGTGATGAACTTGTGTTCAGAGGTCATGAGTGTGATGAT 200
QY 202 GGCATCTTCACTTCACTTCTTTCATCCCAACAGCCCATCTTCACTTCACTTCACT 261
DB 201 GGCATCTTCACTTCACTTCTTTCATCCCAACAGCCCATCTTCACTTCACTTCACT 260
QY 262 TTAATCTGATCTTTAACTTAACTGATGATGATGATGATGATGATGATGATGAT 321
DB 261 TTAATCTGATCTTTAACTTAACTGATGATGATGATGATGATGATGATGATGAT 320
QY 322 ATGGAGAACTTCCAAAGACATTTGATGATGATGATGATGATGATGATGATGAT 381
DB 321 ATGGAGAACTTCCAAAGACATTTGATGATGATGATGATGATGATGATGATGAT 380
QY 382 GTGACATTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441

Db 381 GTGGAAATTAATGACATATTTAGAACTCTGTAAAGATGATTTCAAAATAGAAAGT 440
Qy 442 TATGAGAACTACAGAGTCAAGATTGATGTTGTTCTTGACAGATGCTGTTTTCCCTT 501
Db 441 TATGAAAAAATACACAGAGTCAAGATTTGACATGTTTTGAGATGCTGTTTTCCCTG 500
Qy 502 TGTGAGCTGCTGCGCGAGTTACTTAAATAACCTTTGCTACAGCTCCGCTTCTCC 561
Db 501 TGTGAGCTGCTGCGCTACTTAAATACATCCGTTTGATGAGCTCTCCGCTTCACTCC 560
Qy 562 TGGCTACGCAATTTGAAAAAGCATGTGAGAGACTTGTCTTCCCTCTCTATGTCCTGT 621
Db 561 TGGCTACGCAATTTGAAAAAGCATGTGAGAGACTTGTCTTCCCTCTCTATGTCCTGT 620
Qy 622 TGTATGTCAGAACTAAGTACCAATGACTTCTTACAGAGAGGTTAAAAATATGATCTA 681
Db 621 TGTATGTCAGAAATTAAGTATCAATACTTTCATGAGAGAGGTTAAAAATATGATCTA 680
Qy 682 TGTGCTTATTTTGAATTTGTTCCAAATATTTGACATGAGAGAGTGGATCACTTCTA 741
Db 681 TGTGCTTATTTTGAATTTGTTCCAAATGCTGATGATGAGAGAGTGGATCACTTCTA 740
Qy 742 CAGTGAAGTTCTAGAGAACCCACTACGTTATCTGAGACATGSCAAAAGCTGACATATG 801
Db 741 CAGTGAAGTTCTAGAGAACCCACTACGTTATTTGAGCAATGSGAAAAGCTGACATATG 800
Qy 802 GCTTATTTGAAAATCTAGGAGATTTTCATTTCTCCACCTCAACCAATGTTGAGTT 861
Db 801 GCTTATTTGAAAATCTAGGAGATTTTCATTTCTCCATCATCTTCAACCAAGTTGATTT 860
Qy 862 CGTTGAGAGACTCCACTGCAAACTGCCAAACCCCTACCGAAGAAATGSAAGATTTGT 921
Db 861 TGTGAGAGATTTCCACTGCAAACTGCCAAACCCCTACCTAAGAAATGSAAGATTTGT 920
Qy 922 CCAGAGCTCTGAGAAAAATGTTGTTGTTGTTTCTCTGSGGTGATGTCAGTACAC 981
Db 921 ACAGAGCTCTGAGAAAAATGTTGTTGTTGTTTCTCTGSGGTGATGTCAGTACACAT 980
Qy 982 GTCAAGAAAGGCGCAATGTAAATGTCATGCGCCCTTGCAAGATCCCAAAAGSTTCT 1041
Db 981 GACACAGAAAGGCGCAATGTAAATGTCATGCGCCCTTGCAAGATCCCAAAAGSTTCT 1040
Qy 1042 GTGAGATTTGAGGAAATTAACAGAAATCTTTAGACTCAATACCTGCTGTACAAATG 1101
Db 1041 GTGAGATTTGAGGAAATTAACAGAAATCTTTAGACTCAATACCTGCTGTACAAATG 1100
Qy 1102 GATACCCCAAGATGATCTTTCTGTCATCCAAAACCAAGCTTTTATCACTCATGTGG 1161
Db 1101 GATACCCCAAGATGATCTTTCTGTCATCCAAAACCAAGCTTTTATCACTCATGTGG 1160
Qy 1162 AATGAATGGATCTATGAAGCTATTTACATGAGGCTCCCTATGTTGAGAGTTCCCATATT 1221
Db 1161 AGCCAAATGGATCTATGAAGCTATTTACATGAGGCTCCCTATGTTGAGAGATTTCCCATATT 1220
Qy 1222 TGTGATCAGCTTTGATTAACATAGCTCACATGAAGGCCAAAGAGAGCTGTGAATAAA 1281
Db 1221 TTTGAGATCACTGATTAACATAGCTCACATGAAGGCCAAAGAGAGCTGTGAATAAA 1280
Qy 1282 CTTCAAAAATATAGCAAGGAAATTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1341
Db 1281 CTTCAAAAATATAGCAAGGAAATTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1340
Qy 1342 CTTTATTAAGAAATGCTATGATGATTAATCAAGATTTACCAATGATCAACCTGTAAAGCC 1401
Db 1341 TTTCAATTAAGAAATGCTATGATGATTAATCAAGATTTACCAATGATCAACCTGTAAAGCC 1400
Qy 1402 CTTAGATGAGAGCTTTCTGATGAGATTTGTCTGAGGCCCAAGAGAGCCAAAGCACT 1461
Db 1401 CTTAGATGAGAGCTTTCTGATGAGATTTGTCTGAGGCCCAAGAGAGCCAAAGCACT 1460
Qy 1462 GCGATCAGCTGCCATGACCTGCTGTTCCAGACATACTATAGATGATTTGGGTT 1521
Db 1461 TCGAGTTGAGGCCCAAGCACTGCTGCTGTTCCAGATCACTCTTTGATGATTTGGGTT 1520

Qy 1522 CCTGCTGACCTGTGTGCAACTGCTATATTTCTTTCACAAAAATGTTTTATTTCTCG 1581
Db 1521 CTGCTGCTGTGTGTGCAACTGCTATATTTCTTTCACAAAAATGTTTTATTTCTCG 1580
Qy 1582 TCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1631
Db 1581 CTGAGATTTGCTAGAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 1630

RESULT 8
US-09-981-353-189
; Sequence 189, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Laesk, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 189
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 480489.5
US-09-981-353-189

Query Match 74.7%; Score 1232.4; DB 10; Length 1712;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 231; Indels 1; Gaps 1;

Qy 19 ATTGCATTCATCAGATGTCATGAAATGAGCTTCAGCTTCTCTGATACAGCTGAG 78
Db 1 ATGCATTCACACAGATGATGATGAAATGAGCTTCAGCTTCTCTGATACATCTCAG 60
Qy 79 CTGTTACTTTAGCTCTGAGAGTTGTGAAAGTCTGTGTGCGCCACAGAAATTCAGCA 138
Db 61 TTGTTACTTTAGCTCTGAGAGTTGTGAAAGTCTGTGTGCGCCACAGAAATTCAGCA 120
Qy 139 CTGATGATTAATTAAGCAATCTGATGAACTTGTCCAGAGAGTCAAGTGTGACTGT 198
Db 121 TTGATGATTAATTAAGCAATCTGATGAACTTGTGTCAAGAGTCAAGTGTGACTGT 180
Qy 199 ATTGGCATCTTCAGCTTCATTTCTTCGATCCCAAGAGCCCATCTACTTAAATTTGA 258
Db 181 ACTGGCATCTTCAGCTTCATTTCTTCGATCCCAAGAGCCCATCTACTTAAATTTGA 240
Qy 259 AGTTATCTGTATCTTTACTTAAACTGAGTTGAGATTTATCAAGCAGCTGTTAA 318
Db 241 AGTTATCTGTATCTTTACTTAAACTGAGTTGAGATTTATCAAGCAGCTGTTAA 300
Qy 319 GAGATGGCAGAACTTCCAAAGACACTTTTGGCATATTTTCAAGATTAAGAAAT 378
Db 301 GAGATGGCAGAACTTCCAAAGACACTTTTGGCATATTTTCAAGATTAAGAAAT 360
Qy 379 CATGTGACATTTATGACATATCTAGAAAGTTCTGTAGAGATTAAGTTCAATTAAGAA 438
Db 361 CATGTGACATTTATGACATATTTAGAACTTCTGTAGAAATTAAGTTCAATTAAGAA 420
Qy 439 ACTATGAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGACAGATGCTGTTTCCC 498
Db 421 ACTATGAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGACAGATGCTGTTTCCC 480
Qy 499 CTTTGTGAGCTGCTGCGCGAGTACTTAAATACCCCTTGTACAGCCTCCGCTTCTC 558
Db 481 CTTTGTGAGCTGCTGCGCGAGTACTTAAATACCCCTTGTACAGCCTCCGCTTCTC 540
Qy 559 TCTGAGTACGCAATTAAGAAAGCATAGTGAAGACTTCTGTTCCCTCTCTATGTCGC 618

Db	541	TCCTGGCTACAGAAATTGAAGGCCACAGTGGAGACTGATTTTCCCTCTTCTCACTAC	600
OY	619	TGTTGTTATGTCAGAACTTAAGTGAACCAATGACTTTCAATAGAGGGGTAAAAATATGAT	678
Db	601	TATTGTATTGTCAAAATTAAGTGATCAATGACTTTTCATGGAGAGGGTAAAAATATGAT	660
OY	679	CTATGTGCTTTATTTTGAATTTTGGTCCAAATATTTGACATGAAAGACTGGGATCAGTT	738
Db	661	CTATGTGCTTTATTTTGAATTTTGGTCCAAATATGTCGATATGMAAAGATGGGATCAGTT	720
OY	739	CTACAGTGAAGTTCTAGAAAGCCCACTACGTTATCTGAGCAATGGCGAAAAGCTGACAT	798
Db	721	TTACAGTGAAGTTTTRAGAAAGCCCACTACTTATTTGAGCAATGGGAAAAGCTGACAT	780
OY	799	ATGGCTTATTCGAAACTACGTGGGATTTTCAATTTTCCCAACCACTTTTCCAAATGTTGA	858
Db	781	ATGGCTTATTCGAAAACCTCGGAGTTTCAATTTTCTCAATTCATTTCTTACCAAAAGTTGA	840
OY	859	GTTTCGTGGAGAGATCTCACT - GCAAACTCGCCAAACCCCTACCGAAGAAATGAAAGAT	917
Db	841	TTTTGTGGAGGATTCCTCACTGGCAAACTCGCAAAACCCCTACCTAAGGAAAATGAGAGAT	900
OY	918	TTGTCAGAGCTCTGGAGAAAATGGTGTTGGTGTCTCTCGGGGTGAGATGTCAGTA	977
Db	901	TTGTCAGAGCTCTGGAGAAAATGGTGTTGGTGTCTCTCGGGGTGAGATGTCAGTA	960
OY	978	ACACGTCAGMAAGAAAGGGCCAAATGTAATGCACTACAGCCCTTGCCAAAGTCCCAAAAAG	103
Db	961	ACATGACAGCAAAAAGGGCCAAATGTAATGTCACACAGCCCTTGCCAAAGTCCCAAAAAG	102
OY	1038	TTCTGTGAGATTTGATGSGAAATAAACCAAGATCTTTAGAACTCAATACTCGGCTGTACA	109
Db	1021	TTCTGTGAGATTTGATGSGAAATAAACCAAGATGCTTAAAGTCTCAATACTCGGCTGTACA	108
OY	1098	AGTGAATACCCGAGATGATCTCTTGTGATCCCAAAACCAAGCTTTTATCACTGACAG	115
Db	1081	AGTGAATACCCGAGATGATCTCTTGTGATCCCAAAACCAAGCTTTTATCACTGACAG	114
OY	1158	GTGGAATGAATGGGATCTTATGAAGCTATTACCATGGGGTCCCTTATGTTGGAGTTCCCA	121
Db	1141	GTGGAAGCCAAATGGCATCTTATGAGGCAATCTACCATGGGATCCCTTATGTTGGGACTTCCAT	120
OY	1218	TATTTGGTGATCAGCTTGATAAATAGCTCACATGAAGGCCAAAAGGAGCAGCTGTAGAAA	127
Db	1201	TGTTTGTGATCAACCTGTATAATGCTCACTGAAAGGCCAAAAGGAGCAGCTGTAGAT	126
OY	1278	TAACTTCAAAACTATGACAAAGCAAGATTACTGAGGGCTTTGAGAACATGATTTACCG	133
Db	1261	TGGACTTCAAAACAATGTCGAATGACAGACCTGCTGATCACTGAAGACATTAATTAAG	132
OY	1338	ATTTCCTTAATAAGAAATGCTATGAGATTTTCAAGATTCACCATGATCAACCTGTAA	139
Db	1321	ATCCTTTATATAAGAAATATTAATGAATTTTCAAGATTTCAACATGATCAACAGTAA	138
OY	1398	AGCCCTAGATGAGACAGTCTTCTGGATGAGTTGTCTACATGCGCCCAAAAGGAGCCCAAC	145
Db	1381	AGCCCTAGATGAGACAGTCTTCTGGATGAGTTGTCTACATGCGCCCAAAAGGAGCCCAAC	144
OY	1468	ACCTGCGATCAGCTGCCATGACCTGACTGCTGTTCCAGACATCTCTATATGATGTGATTG	151
Db	1441	ACCTTTCGAGTTGCAAGCCCACTGACCTGCTGTTCCAGTACCACTCTTTGGATGTGATTG	150
OY	1518	GGTTCCTGTCGACCTGTGTGGGAACTGCTAATTTCTGTGTTCAAAAATGTTTTTATTTT	157
Db	1501	GGTTCCTGTCGACCTGTGTGGGAACTGCTAATTTTATCAATCAAAAATGTTTTGTCTGTTT	156
OY	1578	CCTGTCAAAAATTTATATAAATAGTAAGATGAAAAAGGAGATATGATCTTTC	1631
Db	1561	GTTTCGGAAGTTTGTCTAGAAAAAGGAAAGAAAGGAAATTAATGTTATGTC	1614

```

US-10-158-646-42
; Sequence 42, Application US/10158646
; Publication No. US20030073105A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Sornasse, Thierry
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0030-1 US
; CURRENT APPLICATION NUMBER: US/10/158,646
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/295,239
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PERL Program
; SEQ ID NO 42
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030073105A1 480489.3
US-10-158-646-42

Query Match          74.7%; Score 1232.4; DB 14; Length 1712;
Beet Local Similarity 85.6%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 231; Indels 1; Gaps 1

19  ATTGCATTGCATCAGAGTGTCTATGAAATGACATTACGCTCTTCTCCTGATACAGCTGAG 78
Db  1  ATCGGATTGCACAGAGATGACTCTGAAATGACCTTCAGCTTCTTCTGATATCATCTCAG 60

79  CTGTACTTTAGCTCTGAGAGTTGTGAAAAGGTGCTGTGTGAGCCACAGAAATTGACCA 138
Db  61  TTGTTACTTTAGCTCTGGAGAGTTGTGAAAAGGTGCTGTGTGAGCCAGAAATACAGCA 120

139  CTGATGAAATTAAGACAATCCGTGAGTGAATCTGTCCAGAGAGTCATGAGTGACTGT 198
Db  121  TTGGATGAATTAAGAGCAATCTGAAAAGCTTGTCTACAGAGGTCATGAGGTGACTGT 180

199  AATGGCATCTTCAGCTTCCTCATTTCTTGCATCCCAACAGCCCATCTACTTTAAATTTGA 258
Db  181  ACTGGCATCTTCAGCTTCCTCATTTCTTTTGATGCCAATGATGATCCACTTTAAATTTGA 240

259  AGTTTATCCTGATCTTTAACTTAAACCTGAGTTTGAGATATTAATCAAGCAGCTGTTAA 318
Db  241  AGTTTATCCTACATCTTTAACTTAAACCTGAATTTGAGAAATATCATCATGCAACAGGTTAA 300

319  GAGATGGGCAACACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTCAAGAAAT 378
Db  301  GAGATGGTCAGACATTTCCAAAAGATAGCTTTGGTTAATTTTTCACAAGAACAGAAAT 360

379  CATGTCGACATTTAATGACATCTTAGAAGAGTCTGTAAGATATAGTTTCAAATAGAA 438
Db  361  CCGTGGGAAATTAATGACATTTTAGAACTTCTGTAAAGATGTAGTTTCAAATAGAA 420

439  ACTTATGAAGAAACTACAGAGTCAAGATTTGATGTTGTTCTTGAGAGATGCTGTTTCCC 498
Db  421  AGTTATGAAAAAACTACAGAGTCAAGATTTGACATCGTTTGTGAGAGTGTGTTTCCC 480

499  CTTTGTGAGTCTGTGCGCGAGTTACTTAAATATACCCTTTGTCTACAGCTTCGCTTCTC 558
Db  481  CTGTGTGAGTCTGTGCGCGCTTAACTTAACTAGCGTTTGTGTACAGTCTCCGCTTTC 540

559  TCCGTGCTACCAANTGAAAGACATGAGGAGGACCTTGTGTTCCGCTTCCTATGTGCG 618
Db  541  TCCTGTGCTACCAANTGAAAGGACATGAGGAGGACTGATTTTCCCTCTTCTTACATACC 600

619  TGTGTTATGTACAGAACTAAGTGAACCAATGACTTTCATAGAGAGGATPAAAAATATGAT 678
Db  601  TATGTTATGTACAAATTTAAGTGAACCAATGACTTTCATAGAGAGGATPAAAAATATGAT 660

679  CTATGTGCTTATTTGAATTTGGTTCCAAATATTTGACATGAAGAGTGGGATCAGTT 738

```

661 CTATGCTTTATTTTGAAGCTTTGGTTCCTCAATGCTGATATGAAGAAGTGAGTCACTT 720
739 CTACAGGAAGTTCTAGAGAACCCACGATCTTATCTGAGCAATGCAAAAGCTGACAT 798
721 TTACAGGAAGTTTGAAGAACCCACGATCTTATTTAGCAAAAGGGAAGGCTGACAT 780
799 ATGCGTTATTTGAAGAACTCTGAGATTTTCAATTTCTCATCATCTTTACCAAAAGTTGA 858
781 ATGCGTTATTTGAAGAACTCTGAGATTTTCAATTTCTCATCATCTTTACCAAAAGTTGA 840
859 GTTGTGTTGAGAGTCTCACT-6CAAACCTGCCAAACCCCTTACCGAAGAAATGGAAGT 917
841 TTTTGTGAGAGATTTCCACTGGCAAAACCTGCCAAACCCCTTACCGAAGAAATGGAAGT 900
918 TTGTCCAGAGCTCTGAGAAATGAGTGTGTGTGTGTCTTCTGTGGGTGATGTCAGATA 977
901 TTGTACAGAGCTCTGAGAAATGAGTGTGTGTGTGTCTTCTGTGGGTGATGTCAGATA 960
978 ACAGCTCAGAAAGAGGCCAATGTAATTTGATCAGCCCTTGCACAAAGATCCCAAAAG 1037
961 ACATGACAGCAAGAAAGGCCAATGTAATTTGACAGCCCTTGCACAAAGATCCCAAAAG 1020
1038 TTCTGTGAGATTTGATGAGAAATTAACAGATATCTTGAAGTCAATCTCGGCTGTA 1097
1021 TTCTGTGAGATTTGACGGGAATTAACAGATATCTTGAAGTCAATCTCGGCTGTA 1080
1098 AGTGATATCCCAAGATATCTTGTGTGATCCCAAAACCAAGCTTTTATCTCATG 1157
1081 AGTGATATCCCAAGATATCTTGTGTGATCCCAAAACCAAGCTTTTATCTCATG 1140
1158 GTGATATGAAGATCTATGAGCTATTTTACATGAGGCTCTTATGTTGAGGAGTTTCCA 1217
1141 GTGAGCAAGATGATCTATGAGCAATCTTACATGAGGATCCCTATGTTGAGGAGTTTCA 1200
1218 TATTTGTGATGATGATTAATCATAGCTCATAGAGGCCAAGAGCAGCTGTAGAA 1277
1201 TGTTTTGTGATCACTGATTAATCATAGCTCATAGAGGCCAAGAGCAGCTGTAGAT 1260
1278 TAACTTCAAACTATGACAGCAAGATTTATCTGAGGCTTTGAGAACAGCTATTAACG 1337
1261 TGAATTTCAACCAATGTCAGTACAGACCTGCTAAAGCACTGAAGACAGTAATTAATG 1320
1338 ATTCCTTTATTAAGAGATGCTATGATTAATCAAGAAATTCACCATGATCACTGTAA 1397
1321 ATTCCTTTATTAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
1398 AGCCCTAGATGAGCAGCTTTCTGATGATGATGATGATGATGATGATGATGATGATG 1457
1381 AGCCCTAGATGAGCAGCTTTCTGATGATGATGATGATGATGATGATGATGATGATG 1440
1458 ACCCTGATCAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1517
1441 ACCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
1518 GATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1577
1501 GATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
1578 CCTGTCAAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1631
1561 GTTCTGGAAGTTTCTGAAAGAAAGAAAGAAAGAAAGAAAGATTAATGATGATGATG 1614

RESULT 10
US-09-880-107-3756

; Sequence 3756, Application US/09880107
; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3756
LENGTH: 2799
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X63359
US-09-880-107-3756

Query Match 74.3%; Score 1225.2; DB 10; Length 2799;

Best Local Similarity 85.6%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

26 TGCAATCAGATGTTATGAAGATGATCTTCTCTGATACAGCTGAGCTGTAC 85
2 TGCAAGAGATGCTCTGAATGATGATCAG--TTCTGCTGATACAACTGATTTTAC 58
86 TTTAGCTCTGGAGTTGTGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 145
59 TTTAGCTCTGGAGTTGTGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 118
146 AATATGAAGCAATCTGATGAACTTGTCCAGAGAGTCAAGAGTCACTGATTTGCA 205
119 AATATGAAGCAATCTGATGAACTTGTCCAGAGAGTCAAGAGTCACTGATTTGCA 178
206 TCTTCACTTCCATTTCTTTGATCCCAAGCCCATCTCTTAAATTTGAAGTTAT 265
179 TCTTCACTTCCATTTCTTTGATCCCAAGCCCATCTCTTAAATTTGAAGTTAT 238
266 CCTGTATCTTTAACTAAAGCTGATTTGAGATTTATCAAGCAGCTGTTAAGAGATG 325
239 CTTACATCTTTAACTAAAGCTGATTTGAGATTTATCAAGCAGCTGTTAAGAGATG 298
326 GCAGAACTTCCAAAGACATTTTGTGATATTTTTCACAAAGTACAAATCATGTGG 385
299 TCAGAAATTTCAAAAGATATCATTTTGTGATCTTTTTCACAAAGTACAAATCATGTGG 358
386 ACATTTAATGACATCTTAAAGATTTCTGTAAGATTAAGTTTAAATTAAGAACTATG 445
359 GCAATTAATGACATCTTAAAGATTTCTGTAAGATTAAGTTTAAATTAAGAACTATG 418
446 AAGAACTTACAGAGTCAAGATTTGATGTTGCTTTCAGAGTGTGTTTCCCTTGGT 505
419 AAAAACTTACAGAGTCAAGATTTGATGTTGCTTTCAGAGTGTGTTTCCCTTGGT 478
506 GAGTGTGCTGCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 565
479 GAGTGTGCTGCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 538
566 TACGCAATTTGAAGATGATGAGAGCTTCTGTTTCCCTCTCTTCTGATGCTGCTG 625
539 TACTCATTTGAAGATGATGAGAGCTTCTGTTTCCCTCTCTTCTGATGCTGCTG 598
626 ATGTGAGAACTAAGGACCAATGATCTTCAATAGAGAGGTTAAATTAATGATCTATG 685
599 ATGTGAGAACTAAGGATCAATGATCTTCAATAGAGAGGTTAAATTAATGATCTATG 658
686 CTTATTTTGAATTTTGTGTTCAATTAATTTGATGAGAGAGGATGATCTTCAAGT 745
659 CTTATTTTGAATTTTGTGTTCAATTAATTTGATGAGAGAGGATGATCTTCAAGT 718
746 GAAATTTGAGAGAGCCATCTGATCTTATGAGCAATGAGCAAAAGCTGATATGCTT 805
719 GAAATTTGAGAGAGCCATCTGATCTTATGAGCAATGAGCAAAAGCTGATATGCTT 778

```

QY 806 ATTGAAACTACTGGATTTTCAATTTCTCCACCCACTCTTACCAAAATGTTAGTCTGTT 865
   |||||
Db 779 ATGCGAACTCTCGAAATTTTAAATTTCTCATCATTTCTTACCAAAATGTTAGTCTGTT 838
   |||||
QY 866 GAGAGACTCCACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGAGTTGTCCAG 925
   |||||
Db 839 GAGAGACTCCACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGAGTTGTCTAG 898
   |||||
QY 926 AGCTCTGAGAAATGTTGTGTGTGTTCTCTGGGGGTGAGTGTGAGTAAACGTCGA 985
   |||||
Db 899 AGCTCTGAGAAATGTTGTGTGTGTTCTCTGGGGGTGAGTGTGAGTAAACGTCGA 958
   |||||
QY 986 GAGAAAGGCGCAATGTAATTCATCAGCCCTTCGCAAGATCCCAAAAAGTTCTGTGG 1045
   |||||
Db 959 GAAAGAAAGGCGCAATGTAATTCATCAGCCCTTCGCAAGATCCCAAAAAGTTCTGTGG 1018
   |||||
QY 1046 AGATTTGATGGGAATTAACCAAGATCTTTAGAGACTCAATCTCGGCTGTACAGTGATA 1105
   |||||
Db 1019 AGATTTGATGGGAATTAACCAAGATCTTTAGAGACTCAATCTCGGCTGTACAGTGATA 1078
   |||||
QY 1106 CCCGAGATGATCTTTGTGTCATCCCAAAACCAAGCTTTATCACTCATGTGGAAATG 1165
   |||||
Db 1079 CCCGAGATGATCTTTGTGTCATCCCAAAACCAAGCTTTATCACTCATGTGGAAATG 1138
   |||||
QY 1166 AATGGATCTATGAGCTAATTTACCATGGGGTCCCTAATGTGGAGATTCCTCATATTTGGT 1225
   |||||
Db 1139 AATGGATCTATGAGCTAATTTACCATGGGGTCCCTAATGTGGAGATTCCTCATATTTGGT 1198
   |||||
QY 1226 GATCAGCTTGATTAATAGCTCACAATGAGGCGCAAGAGAGAGCTGTAGAAATTAACCTTC 1285
   |||||
Db 1199 GATCAGCTTGATTAATAGCTCACAATGAGGCGCAAGAGAGAGCTGTAGAGTGAAGCTTC 1258
   |||||
QY 1286 AAAACTATGACAAACGAAGATTTACTGAGGGCTTTGAGAAAGATCATTCAGATTCCTCT 1345
   |||||
Db 1259 AAACCAATGTCAGATTAACGAGCTCTGATGATGAGACAGATTAATTAATGATCTTCA 1318
   |||||
QY 1346 TATTAAGAGATGCTATGAGATTAATCAAAATTCACATGATCAACCTGTAAAGCCCTTA 1405
   |||||
Db 1319 TATTAAGAGATTAATTAATTAATTAATCAAGAAATTCAGATGATCAACCTGTAAAGCCCTTA 1378
   |||||
QY 1406 GATGAGCAGCTCTTCTGATGAGATTTGTCTCATGCGCCCAAAAGAGCGCAACCTGTGGA 1465
   |||||
Db 1379 GATGAGCAGCTCTTCTGATGAGATTTGTCTCATGCGCCCAAAAGAGCGCAACCTGTGGA 1438
   |||||
QY 1466 TCAGCTGCGCCATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1525
   |||||
Db 1439 GTTGCAAGCCCAACCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1498
   |||||
QY 1526 CTGACCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCTGTCGAA 1585
   |||||
Db 1499 CTGACCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCTGTCGAA 1558
   |||||
QY 1586 AAATTTAATAAACTAGAAAGATGAAAAAGAGGAAATGATCTTTC 1631
   |||||
Db 1559 AAGTTTCTAGAAAAAGAAAGAAAGAAAGAAAGGATTAATATATC 1604
   |||||

```

```

; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: InCyte ID No. US20020160382A1 255115.4
;
; NAME/KEY: unsure
; LOCATION: 2087, 2089, 2094, 2096-2098, 2108, 2110, 2112, 2115-2116, 2120,
; LOCATION: 2122-2123, 2125, 2136
; OTHER INFORMATION: a, t, c, g, or other
US-09-981-353-45

Query Match      68.9%; Score 1137.2; DB 10; Length 2150;
Best Local Similarity 82.1%; Pred. No. 6.5e-306;
Matches 1345; Conservative 0; Mismatches 288; Indels 5; Gaps 3;

QY 1 AGCAACTGGAAGAAACAAGCATTTGATTCATGAGATGTCATATGAAATGAGCTTACGCTCT 60
   |||||
Db 18 AACCACTGGAAGAAAGAAACATTTGATTCATGAGATGTCATATGAAATGAGCTTACGCTCT 77
   |||||
QY 61 TCTCCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGGAAGGTGCTGTGTG 120
   |||||
Db 78 TCTCCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGGAAGGTGCTGTGTG 137
   |||||
QY 121 GCCCAGAAATTAAGCCATGATGAATATPAAGACAAATCTGAGTAACCTGTCAGAG 180
   |||||
Db 138 GCCCAGAAATTAAGCCATGATGAATATPAAGACAAATCTGAGTAACCTGTCAGAG 197
   |||||
QY 181 AGGTCAAGAGTACATGATTTGGATCTTCACTTCTTTGATCCCAACAGCC 240
   |||||
Db 198 GGGTCAGAGAGTATGTTGTGATCTTGGCTTCACTTCTTTGATCCCAACAGCC 257
   |||||
QY 241 ATCTACTCTTAAATTTGAAATTTATCTGATCTTTACTTAACTGAGTTTGAGATAT 300
   |||||
Db 258 ATCTGCTATTAATTAAGAAATTTATCTGATCTTTACTTAACTGAGTTTGAGATAT 317
   |||||
QY 301 TATCAAGCAGCTGTTAAGAGATGGGAGA--ACTTCAAAAAGACATTTTGTGTCATA 357
   |||||
Db 318 TTTTATGAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 377
   |||||
QY 358 TTTTTCACAAAGTACAAAGAAATCATGTGAGACATTTAATGACATCTTGAAAGTTCTGTAA 417
   |||||
Db 378 TTTTTCACAAAGTACAAAGAAATCATGTGAGATTTCTGACTATTAATTAAGCTGTGTGA 437
   |||||
QY 418 GGAATATGTTTCAATTAAGAACTTATGAGAAACCTACAGAGCTCAAGATTTTGTGATGTTGT 477
   |||||
Db 438 AGATGACGTTTGAACAAAGAACTTATGAGAAACCTACAGAGCTCAAGATTTTGTGATGTTGT 497
   |||||
QY 478 TCTTGCAAGATGCTGTTTCCCTTTGTGAGAGCTGTCGCGAGTTACTTAAATACCTT 537
   |||||
Db 498 TCTTGCAAGATGCTGTTTCCCTTTGTGAGAGCTGTCGCGAGTTACTTAAATACCTT 557
   |||||
QY 538 TGTCTACAGCTTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 597
   |||||
Db 558 TGTCTACAGCTTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 617
   |||||
QY 598 GTTCCCTCTCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
   |||||
Db 618 GTTCCCTCTCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 677
   |||||
QY 658 AGAGAGGTAATAAATATGATGCTATGCTTTATTTTGAATTTTGTGTTCCAAATATTTGA 717
   |||||
Db 678 GAGAGAGTAAATAAATATGATGCTATGCTTTATTTTGAATTTTGTGTTCCAAATATTTGA 737
   |||||
QY 718 CATGAGAAATGGGATGATGTTCTACAGTGAAGTTCTAGAAAGACCTACATGTTATCTGA 777
   |||||
Db 738 TCTGAAGAAATGGGATGATGTTCTACAGTGAAGTTCTAGAAAGACCTACATGTTATTTGA 797
   |||||
QY 778 GACAATGGCAAAAGCTGACATATGCTTTATTCGAAATCTACGAGATTTTCAATTTCTCTCA 837
   |||||
Db 798 GACAATGGGAAAGCTGAAATGCTCATTTGCAACTATTTGGGATTTTGAATTTCTCTCG 857
   |||||
QY 838 CCCACTCTTACCAAAATGTTAGTTCGTTGGAGAGCTCACTGCAAAACCTGCAAAACCCCT 897
   |||||
Db 858 CCCACTCTTACCAAAATGTTAGTTCGTTGGAGAGCTCACTGTAACCAAGCAAAACCCCTT 917
   |||||

```

```

RESULT 11
US-09-981-353-45
; Sequence 45, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Laeek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981.353
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 2150
; TYPE: DNA

```

```

Qy 898 ACCGAGGAATGGAAGAGTTTGTCCAGAGCTCTGGAGAAATGGTGTGTGTGTTTC 957
Db 918 GCTTAGGAATGGAAGAGTTTGTCCAGAGCTCTGGAGAAATGGTGTGTGTTTC 977
Qy 958 TCTGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1017
Db 978 TCTGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1035
Qy 1018 TGCCAAATGCCCAAAAGGTTCTGTGAGATTTATGGAATTAACCAATGTTTGG 1077
Db 1036 TGCCAAATGCCCAAAAGGTTCTGTGAGATTTATGGAATTAACCAATGTTTGG 1095
Qy 1078 ACTCAATCTGGGCTGTGCAAGTGTGATCCCAAGATGATCTTGTGTGTGTGTGTGT 1137
Db 1096 TTTCAATCTGTGCTGTGCAAGTGTGATCCCAAGATGATCTTGTGTGTGTGTGTGT 1155
Qy 1138 CAAGGCTTTTATCACTCATGTGTGAAATGGAATGGAATGGAATGGAATGGAATG 1197
Db 1156 CAAGGCTTTTATCACTCATGTGTGAAATGGAATGGAATGGAATGGAATGGAATG 1215
Qy 1198 CCTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1257
Db 1216 CCTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1275
Qy 1258 CAAGGAGGAGCTGTGAAATTAATCAAACTATGA CAAGGAGATTTACTGAGGGC 1317
Db 1276 CAAGGAGGAGCTGTGAAATTAATCAAACTATGA CAAGGAGATTTACTGAGGGC 1335
Qy 1318 TTTGAGAAAGTATTCAGATTCCTCTTATTAAGAGAATGCTATGATTAATCAAGAT 1377
Db 1336 ATTTGAGTATGATTAATTAAGAGCTTATCTATTAAGAGAATTAATCAAGATTAATCA 1395
Qy 1378 TCACATGATCACTGTGAAGCCCTTGTGATGTGAGAGTCTTGTGTGTGTGTGTGTGT 1437
Db 1396 TCACATGATCACTGTGAAGCCCTTGTGATGTGAGAGTCTTGTGTGTGTGTGTGTGT 1455
Qy 1438 GCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1497
Db 1456 GCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1515
Qy 1498 CTACTCTATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1557
Db 1516 CCACCTTTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1575
Qy 1558 CACAAATGTTTTTATTTTCTGTCAAAATTTAATAAATAAGATTAAGAAAGAG 1617
Db 1576 CACAAATGTTTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1635
Qy 1618 GGAATGATCTTTCCAA 1635
Db 1636 GGAATGATCTTTCCAA 1653

```

```

RESULT 12
US-10-252-157-25
; Sequence 25, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Pearle, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 25
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 25115.4
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2087, 2089, 2094, 2096-2098, 2108, 2110, 2112, 2115-2116, 2120, 2122-2123,
; LOCATION: 2125, 2136
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-25

```

```

Query Match 68.9%; Score 1137.2; DB 12; Length 2150;
Best Local Similarity 82.1%; Pred. No. 6.5e-306;
Matches 1345; Conservative 0; Mismatches 288; Indels 5; Gaps 3;

```

```

Qy 1 AGCAATGGAAGAAAGAGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 18 AACAACTGGAAGAAAGAGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 77
Qy 61 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 78 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 137
Qy 121 GCCCAAGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
Db 138 GCCCAAGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 197
Qy 181 AGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db 198 AGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 257
Qy 241 ATCTACTCTTAAATTTGAGGATTTATCTGTATCTTTAATCTTAAATCTGAGTTGAGATAT 300
Db 258 ATCTACTCTTAAATTTGAGGATTTATCTGTATCTTTAATCTTAAATCTGAGTTGAGATAT 317
Qy 301 TATCAAGAGGAGGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 357
Db 318 TTTTATGAAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377
Qy 358 TTTTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 417
Db 378 TTTTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 437
Qy 418 GGAATATGATTTCAATATGAAGAACTTATGAGAGAACTACAGAGTCAAGATTTGATGTGT 477
Db 438 AGATGAGGATTTGAGAGGAACTTATGAGAGAACTACAGAGTCAAGATTTGATGTGTGT 497
Qy 478 TCTTGCAGATGCTGTTTTTCCCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 537
Db 498 TCTTGCAGATGCTGTTTTTCCCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 557
Qy 538 TGTCTACAGCTCCGCTTCTCTCTGCTACGCAATGGAAGATGAGGAGGAGGAGGAGGAGG 597
Db 558 TGTCTACAGCTCCGCTTCTCTCTGCTACGCAATGGAAGATGAGGAGGAGGAGGAGGAGG 617
Qy 598 GTTCCCTCTCTCTATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 657
Db 618 GTTCCCTCTCTCTATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 677
Qy 658 AGAGAGGATTAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
Db 678 AGAGAGGATTAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 737
Qy 718 CATGAAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
Db 738 TCTGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 797
Qy 778 GACAAATGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 837
Db 798 GACAAATGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 857
Qy 838 CCAATCTTAACTAAATGTTGATGTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 897
Db 858 CCAATCTTAACTAAATGTTGATGTTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 917

```


OY		898	ACCGAAGAAATGGAAGAGTTTGTGCAGAGCTCTGGAGAATAATGGTGTTGGTCCTTTC	957
Db		918	GCCTAGAAGAAATGGAAGAGTTTGTGCAGAGCTCTGGAGAATAATGGTGTTGGTGTTTTTC	977
OY		958	TCTGGGGTCGATGTGTCAGTAACAACGTCAGAAAGAAAGGGCCAATGTAAATTGCATCAGCCCT	1017
Db		978	TCTGGGGTCGATGTGATCACTAACAATGTCAGAAAGAAAGTGCCAAATGTA-TGCATCAGCCC-	1033
OY		1018	TGCCAAGATCCCAAAAAGGTTCTGTGAGATTTTATGSGAAATTAACCAATATCTTTTGG	1077
Db		1036	TGCCCAAGATCCCAAAAAGGTTCTATGAGATTTTATGSCAAGAAAGCCAAATATCTTTTGG	1095
OY		1078	ACTCAATCTCGGCTGTGAACAAGTGAATACCCAGAAATGATCTCTTGGCATCCCAAAAC	1133
Db		1096	TTCCAAATCTCGACTGTCAAGATGTTATACCAGAAATGATCTTTTGGTCAATCCAAAAAC	1155
OY		1138	CAAAGCTTTTATCTACTCATGTGTGAATGAATGGATCTATGAAAGCTATTTAACATGGGGT	1197
Db		1156	CAAAGCTTTTATTAATCTCATGTGTGAACCAATGGCATCTATGAGGCAATCTACATGGGAT	1215
OY		1198	CCCTATGGTGGAGATTTCCCATATTTTGGTGATCAGCTTGATTAACATAGCTCACATGAAGC	1257
Db		1216	CCCTATGGTGGAGATTTCCCTGTTTGGAGATCAATGATATAATGCTCAATGAAAGC	1275
OY		1258	CAAAGAGACAGCTGTAGAATAATTAACCTTGAAACCTATGA CAAGCAAGATTTACTGAGGGC	1317
Db		1276	CAAGGAGACAGCCCTCAGTGTGACATAGAGACATGTCAAGTAGAATTTTGCTCAATGC	1335
OY		1318	TTTGAAGAACAGTCATTACCGAATCTCTCTTATTAAGAAGATGCTATGATTAATCAAGAT	1377
Db		1336	ATTGAAGTCAGTCATTTAATGAACCCATCTATCTATTAAGAAATATCATGAATTAATCAAGAT	1395
OY		1378	TCACCATGATCAACCTGTAAAGCCCCCTTGATTCGACAGCTCTTTGGATTCGAGTTTGAT	1433
Db		1396	TCATCATGATCAACCGGTGAAGCCCCCTTGATTCGACAGCTCTTTGGATTTGAGTTTGAT	1455
OY		1438	GGCGCACAAAGAGGCAAGACCTGCGATCAGCTGCCATGCCATGCACTCACTGGTTCACGA	1497
Db		1456	GGCGCATTAAGAGAGCAAGACCTTCGGGTGGACAGCCCAACACCTGACCTGATTCAGTA	1515
OY		1498	CTACTCTATGATGTGATVGGGTTCTGTGCTGACCTGTGGCAACTGCTATATCTTGTT	1557
Db		1516	CCACTCTTTGGATGTGATGACATTCCTGCTGAGCCGTGGCAACTGATATATTTATGAT	1575
OY		1558	CACAAAAGTTTATTTATTTTCTGTGCAAAAATTTATTAATCTAGAAAAGATAGAAAAG	1617
Db		1576	CACAAAAGTTTATTTTCTGTGTTTCTGCAAAAAGTTGCAAAAACAGAAAAGAAAAG	1635
OY		1618	GGAATAGATCTTTCCAA 1635	
Db		1636	GGATTAGTTATATCAAA 1653	
<hr/>				
RESULT 13				
US-10-252-157-24				
Sequence 24, Application US/10252157				
Publication No. US20030190640n1				
GENERAL INFORMATION:				
APPLICANT: Paris, Mary				
APPLICANT: Pearson, Cecelia I.				
FILE REFERENCE: PA-0027-1 US				
CURRENT APPLICATION NUMBER: US/10/252,157				
PRIOR FILING DATE: 2002-10-01				
PRIOR APPLICATION NUMBER: 60/295,048				
NUMBER OF SEQ ID NOS: 501				
SOFTWARE: PERL Program				
SEQ ID NO 24				
LENGTH: 1829				
TYPE: DNA				
ORGANISM: Homo sapiens				

```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER_INFORMATION: Incyte ID No. US20030190640A1 255115.2
US-10-252-157-24

```

Query Match	68.8%;	Score 1135.8;	DB 12;	Length 1829;
Best Local Similarity	-81.6%;	Pred. No. 1.4e-305;		
Matches 1327;	Conservative 0;	Mismatches 297;	Indels 3;	Gaps 1;

OY	12	AACAAGCATTCGATCGATCAGAGATGTCTAAGAAATGAGACTCAGCTCTTCTCCGATAC	71
Db	2	AAGAAGCATTCGATTAAGACCAAGATGTCTCGAAATGAGACGCGACTCTTCTCGTGATAC	61
OY	72	AGCTGAGCTGTAACCTTAGCTCTGGAGATGTGAAAGGTGCTGGTGTGGCCACAGAT	131
Db	62	AGCTCAGTGTATCTTAGCTCTGGAACTGTGAAAGGTGTGATGTGGCCACAGAT	121
OY	132	TCAGCCACTGGAATGAATTAAGAACAATCCCTGATGAACTTGTCAGAGAGTCAATGAG	191
Db	122	ACAGCCATTTGATTAATATGAAAGCAATCCGGAAGAGCTTGTCAGAGAGTCAATGAG	181
OY	192	TGACTGTAATGGACCTTCAGCTCCATTTCTTGATGCCAACGCCCATCTACTCTTA	251
Db	182	TGACTGTAATGGACCTTCAGCTCCATTTCTTGATGCCAACGCCCATCTACTCTTA	241
OY	252	AATTGAAGTTATCTGTATCTTTAACTGAAACTGAGTTTGAGATATTAACAAGAC	311
Db	242	AATTGAAGTTATCTGTATCTTTAACTGAAACTGAGTTTGAGATATTAACAAGAC	301
OY	312	TGCTTAAGAGATGGGCA---GAACCTCCAAAAGACACATTTTGTCATATTTTACAAG	368
Db	302	TTCTCGATAGATGATATATGTGTGTCMAAAATACATTTTGTCATATTTTACAAT	361
OY	369	TACAAGAAATCATGGACATTTATATGACATATCTAGAAAGTTCGTGAAGATATAGTT	428
Db	362	TACAAGAAATCATGGACATTTATATGACATATCTAGAAAGTTCGTGAAGATATAGTT	421
OY	429	CAAAATAGAAACTTATGAGAAAGATACAGAGATCAAGATTTGATGTGTTCTTGACAGATG	488
Db	422	TGAATTAAGAACTTATGATGAAAGATACAGAGATCAAGATTTGATGTGTTCTTGACAGATG	481
OY	489	CTGTTTTCCCTTTGGTGAAGTCTGTGGCCGATTACTTAAATVACCCTTTGTCTACAGCC	548
Db	482	CCCTTAATCCCTGTGGAGCTACTGGGTGAACATTTAAATVACCCTTTGTCTACAGCTC	541
OY	549	TCCGCTCTCTCCCGGCTACGCACTTGAAGAAGCATAGTGGAGGACTTGTCTCCCTCTT	608
Db	542	TTGATTTCTGTGGCTACCACTTTGAAGAAGATGTGGAGGATTTGTCTCCCTCTT	601
OY	609	CCATATGCGCTGTGTATATGCACAAGCTAACTGACCAATACCTTTCTATGAGAGGATTA	668
Db	602	CCATATGACCTGTGTATATGCACAATTAAGTATCAAAATGATTTCTATGAGAGGATTA	661
OY	669	AAAAATATGATCTATGTGCTTTATTTTGAATTTGGTTCMAAATTTTGACATGAAGAAT	728
Db	662	AAAAATATGATATATATGCTTTATTTTGAATTTGGTTCMAAATTTTGACATGAAGAAT	721
OY	729	GGGATCAGTTCTACAGTGAAGTTCTTAGAAGAACCCACTAGCTATCTGAGACATGSCAA	788
Db	722	GGGACCAAGTTTATATGAGAGTTCTTAGAAGAACCCACTAGCTATCTGAGACATGSCAA	781
OY	789	AAGCTGACATATGGCTTAATCGAAAGCTACTGGAGATTTTCATTTTCTGACCACATCTTAC	848
Db	782	AAGCTGAAATGTGCTCATTCGAACTTAATGGAGATTTTGAATTTCTGCCCCATTTCTAC	841
OY	849	CAAAATGTGAAGTTGTTGAGAGACTCACTGCAAACTTGCCAAAACCCCTACCGAAGAAA	908
Db	842	CAAAATGTGAATTTGTTGAGAGACTTCACTGTAAACCAAGCCAAACCCCTGCTAAGAAA	901
OY	909	TGGAAGATTTGTCCANAGCTCTGGAGAAAAATGTGTGTGGTGTGTTTCTCTGGGGTGA	968
Db	902	TGGAAGATTTGTCCANAGCTCTGGAGAAAAATGTGTATTTGGTGTGTTTCTCTGGGGTGA	961

```

OY 969 TGGTCAGTAACACGTCAGAGAAAGGCCAATGTAATGATCAGCCCTTGCCAGATCC 1028
DB 962 TGATCAGTAACATGTCAGAGAAAGTGCACATGATTCAGCCCTTGCCAGATCC 1021
OY 1029 CACAAAAGGTTCTGTGAGATTTGATGGGAATAAACAGATACCTTAGACTCAATATC 1088
DB 1022 CACAAAAGGTTCTGTGAGATTTGATGGGAATAAACAGATACCTTAGACTCAATATC 1081
OY 1089 GGGCTGTACAGTGTATACCCAGAAATGATCTTTCTTGTCATCCCAAAACCAAGCTTTTA 1148
DB 1082 GACTGTACAGTGTATACCCAGAAATGATCTTTCTTGTCATCCCAAAACCAAGCTTTTA 1141
OY 1149 TCATCAGTGTGAGATGAGATGAGATCTATGAGATTAATTAACATGGGGTCCCTATGATG 1208
DB 1142 TAACTCAGTGTGAGATGAGATGAGATCTATGAGATTAATTAACATGGGGTCCCTATGATG 1201
OY 1209 GAGTTCCCATATTTGTGATCAGCTTGATTAATAGCTCAGATGAAAGGCCAAAGAGCAG 1268
DB 1202 GCATTCCTGTTGGCCGATCAACATGATTAATGCTCAGATGAAAGGCCAAAGAGCAG 1261
OY 1269 CTGTAGAAATTAACCTTCAAACTATGACAGCGAAAGTTTACTGAGGGCTTTGAAACAG 1328
DB 1262 CCTCAGTGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1321
OY 1329 TCATTAACGATTCCTCTATTAAGAGATGCTATGAGATTAATTAACATGAGATGAGATG 1388
DB 1332 TCATTAAGACCTGCTCTATTAAGAGATGCTATGAGATTAATTAACATGAGATGAGATG 1381
OY 1389 AACCTGTAAGCCCTAGATGAGAGATCTTCTGATGAGATGAGATGAGATGAGATGAG 1448
DB 1382 AACCAATGAGCCCTGATGAGAGATCTTCTGATGAGATGAGATGAGATGAGATGAGATG 1441
OY 1449 GAGCCAGACCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1508
DB 1442 GAGCCAGACCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1501
OY 1509 ATGATGAGGATCTCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1568
DB 1502 ATGATGAGATCTCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1561
OY 1569 TTTTATTTTCTGTGCAAAATTTAATAAATGAGAGATGAGAGAGAGATGATCT 1628
DB 1562 GCGTGTGTTGTTCCGAAAGCTTGCAAAACAGAGAGAGAGAGATGATGATCT 1621
OY 1629 TTCCAAA 1635
DB 1622 ATCAAAA 1628

```

```

; NAME/KEY: CDS
; LOCATION: (11) ... (1598)
US-10-205-522-112

Query Match 68.3%; Score 1126.8; DB 14; Length 1976;
Best Local Similarity 81.7%; Pred. No. 4,9e-303;
Matches 1315; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

OY 29 ATCAGATGTCATGAGAAATGAGATGAGATCTTCTCTGATATACAGTGTGATCTT 88
DB 5 ACCAGATGTCATGAGAAATGAGATGAGATCTTCTCTGATATACAGTGTGATCTT 64
OY 89 AGCTGTGAGATGAGAAAGTGTGAGTGTGAGCCCAAGAAATGAGCCATGAGAT 148
DB 65 AGCTGTGAGATGAGAAAGTGTGAGTGTGAGCCCAAGAAATGAGCCATGAGAT 124
OY 149 ATTAAGCAATCTGATGAACTTGTCCAGAGAGATATAGAGTGTGATGATGATCT 208
DB 125 ATGAAGCAATCTGAGAAAGTGTGATGAGAGGATATAGAGTGTGATGATGATCT 184
OY 209 TCAGCTCCATTTCTTGATGCCCAAGCCATCTACTTAAATTTGAAATTTATCT 268
DB 185 TCGCTTCTACTCTTGTCAATGCCAGTAATCATCTCTATTAATTTGAAATTTATCT 244
OY 269 GTATCTTAACTAACTGATTTGAGATATTAACAAGAGCTGTAAAGATGAGCA 328
DB 245 ACATCTTAACTAAATGATTTGAGATTTCTCTGAAATTTCTGATAGATGATA 304
OY 329 ---GAACTTCAAAAGACATTTTGTGATTTTTCACAGTACAAAGAAATGATG 385
DB 305 TATGATTTCAAAATTAATTAATTTTGTGATTTTTCACAGTACAAAGAAATGATG 364
OY 386 ACATTAATGATATTAAGTGTGATGAGATTAATGATTTCAATTAAGAAATTAATG 445
DB 365 GAATTTATGATCTACAGTACAGTGTGATGAGATGAGATTTGATTAAGAAATTAATG 424
OY 446 AAGAACTACAGAGATGAGATTTGATGATGTTGTTTTCAGATGCTGTTTCCCTTGT 505
DB 425 ATGAATCTACAAAGATCAAGATTTGATGATGATTTCTGAGATGCTTAAATCCCTG 484
OY 506 GAGCTGCTGCGGATTAATTTAAATACCTTTGTCTACAGCTCCGCTTCTCTGAGC 565
DB 485 GAGCTGCTGCGGATTAATTTAAATACCTTTGTCTACAGCTCCGCTTCTCTGAGC 544
OY 566 TACGCAATGAAAGCAATGAGAGAGATCTGTTTCCCTCTCTTATGCTGTTGT 625
DB 545 TACGCAATGAAAGCAATGAGAGAGATCTGTTTCCCTCTCTTATGCTGTTGT 604
OY 626 ATGTCAGAACTAGTGAACCAATGATCTTCAATGAGAGAGATTAATGATCATGTG 685
DB 605 ATGTCAGAACTAGTGAACCAATGATCTTCAATGAGAGAGATTAATGATCATGTG 664
OY 686 CTTTATTTGATTTTGTGTTCCAAATATTTGACATGAAAGTGTGATCTTCTACAGT 745
DB 665 CTTTATTTGATTTTGTGTTCCAAATATTTGATGATGAAAGTGTGATCTTCTACAGT 724
OY 746 GAGTTCTAGAGAGAGCCACTAGCTTATCTGAGCAATGAGCAAAAGCTGATAGCTT 805
DB 725 GAGTTCTAGAGAGAGCCACTAGCTTATTTAGAGCAATGAGCAAAAGCTGATAGCTT 784
OY 806 ATTGCAAACTAGTGAATTTTCAATTTCTTCAACCACTCTTACCAATGTTGATGCTT 865
DB 785 ATTGCAAACTAGTGAATTTTCAATTTCTTCAACCACTCTTACCAATGTTGATGCTT 844
OY 866 GAGAGACTCACTGCAAACTGCAAAACCTTACCGAAGAAATGAAAGATTTGTCCAG 925
DB 845 GAGAGACTCACTGCAAACTGCAAAACCTTACCGAAGAAATGAAAGATTTGTCCAG 904
OY 926 AGCTCTGAGAAATTTGTTGTTGTTTCTGAGGAGTGTGATGATGATGATGATGATG 985
DB 905 AGCTCTGAGAAATTTGTTGTTGTTTCTGAGGAGTGTGATGATGATGATGATGATG 964
OY 986 GAAGAAAGGCCAATGTAATGATGATGATGATGATGATGATGATGATGATGATG 1045

```


Db 965 GAAGAAAGTGCACATGATGTCAGCCCTTGCCCGATGCCACAAAGGTTCTATGG 1024
Qy 1046 AGATTTGATGGGATTAACCGAGATCTTAGAGCTCAATCTCGGCTGTAAAGTGATA 1105
Db 1025 AGATTTGATGGCAAGAACCAAAATCTTAGGTTCCAACTGCTGTAAAGTGATA 1084
Qy 1106 CCCGAGATGATCTTCTGTGTCATCCCAAAACCAAGCTTTATCACTCATGCGGATG 1165
Db 1085 CCCGAGATGATCTTCTGTGTCATCCCAAAACCAAGCTTTATCACTCATGCGGATG 1144
Qy 1166 AATGGATCTATGAGATTAATTACCATGAGGTCCTCATGATGAGGATTTCCATATTTGGT 1225
Db 1145 AATGGATCTATGAGGATCTACATGATGATCCCTATGATGAGGATTTCCCTTTGTTGG 1204
Qy 1226 GATCAGCTTATATCATGCTCATCATGAGGCCAAAGAGACAGCTGTAAATTAATCTTC 1285
Db 1205 GATCAGATGATTAATCATTTGCTCATGTAAGCCAAAGGAGAGCCCTCATGATGAGCATC 1264
Qy 1286 AAAAATATGACAGGAGATTTTACTGAGGCTTTGAGAACAGTCAATTAACGATTTCCCT 1345
Db 1265 AGGACATGATGATGAGATTTTGTCTCATGATGATGAGTCAATTAATTAAGCCCTGTC 1324
Qy 1346 TATAAGAGAAATGCTATGATTAATCAAGAAATTCACATGATCAACCTGTAAGCCCTTA 1405
Db 1325 TATAAGAGATGCTATGATTAATCAAGAAATTCATGATCAACCAATGAAGCCCTG 1384
Qy 1406 GATGAGAGATCTTCTGTGATGAGTTTGTCTATGCGCCCAAAAGAGCCAGACCTGCGA 1465
Db 1385 GATGAGAGATCTTCTGTGATGAGTTTGTCTATGCGCCCAAAAGAGCCAGACCTTCTGA 1444
Qy 1466 TCAGCTGCCATGACCTCAGCTGTTCCAGCATCTATAGATGATGAGGTTCCCTG 1525
Db 1445 CTCGAGCTGCACAACTCAGCTGATCCAGTACAGTACATCTTTGATGATGATGATCTG 1504
Qy 1526 CTGACCTGTGTGGAACCTGCTATATCTTGTGTCAAAATGTTTTTATTTCCGTGCA 1585
Db 1505 CTGCGCTGCGTGGAACCTGCTATATCTTGTGTCAAAATTTTGTGCTTTGTTCCGA 1564
Qy 1586 AAATTTAATAAATACTAGAAAGATGAAAAAGGAGATGATCTTCCAA 1635
Db 1565 AAGCTTGCCAAACAGGAAAGAAAGAAAGATGATTAATCAAAA 1614

RESULT 15
US-09-880-107-3292
; Sequence 3292, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3292
; LENGTH: 2090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U08854
US-09-880-107-3292

Query Match 68.2%; Score 1125.2; DB 10; Length 2090;
Best Local Similarity 81.6%; Pred. No. 1.4e-302;

Matches 1314; Conservative 0; Mismatches 293; Indels 3; Gaps 1;
Qy 29 ATGAGATGCTATGAAATGAGACTTACGCTCTCTCTGTATPACAGCTGTTACTTT 88
Db 16 ACCAGATGCTCTGAAATGAGAGTCACTCTTCTGTATACAGCTGATTTACTTT 75
Qy 89 AGCTCTGGAGTTGTGGAAGAGTCTGCTGTCGCCACAGAAATTCAGCCACTGATGAAT 148
Db 76 AGCTCTGGAAGCTGTGGAAGAGTCTGCTGTCGCCACAGAAATTCAGCCACTGATGAAT 135
Qy 149 ATAAAGCAATCTCGATGAATCTTGTCTCAGAGAGCTATGAGTGAATGATGCTAT 208
Db 136 ATGAAGCAATCTCGAAGAGCTTGTCTCAGAGAGCTATGAGTGAATGATGCTAT 195
Qy 209 TCAGCTTCAATTTCTTGTGATCCCAAGCCCAATCTCTTAATTTGAATTTATCT 268
Db 196 TCGCTTCTACTCTTGTCAATGCGAGTAAATCATCTGCTATTAATTAAGAGTTATCT 255
Qy 269 GTATCTTAACTAAATCGATTTGAGATATTAATCAAGAGCTGTTAAGAGATGGCA 328
Db 256 ACATCTTTAATAAATGATTTGGAAGATTTCTCTGAAATTTCTCGATAGATGATA 315
Qy 329 - - - GAACCTTCAAAAGACATTTTGTGCTATATTTTTCACAGTCAAGAAATCATG 385
Db 316 TATGCTTTTCAAAATTAATATTTTGTGCTATATTTTTCACAGTCAAGAAATGTTG 375
Qy 386 ACATTTAATACACTTATGAAATGTTCTGTAAGATTAATTTCAATTAAGAACTATG 445
Db 376 GAATTTATATACATACATGTAACAGCTGTAAAGATGCAATTTGAAGAACTATG 435
Qy 446 AAGAACTAAGAGAGTCAAGATTTGATGTTGTTCTTTCAGATGCTGTTTTCCCTTTGT 505
Db 436 ATGAATCTAAGAGTCAAGATTTGATGTTGTTCTTTCAGATGCTGTTTTCCCTTTGT 495
Qy 506 GAGCTGTGCGGAGTTACTTAATAATCCCTTGTCTACAGCTCGCTTCTCTCTGCG 565
Db 496 GAGCTGTGCGGAGTTACTTAATAATCCCTTGTCTACAGCTCGCTTCTCTCTGCG 555
Qy 566 TACGCAATTAAGAGTCTATGAGAGCTTCTGTTCCCTCTCTGATGAGCTGTTGT 625
Db 556 TACGCAATTAAGAGTCTATGAGAGCTTCTGTTCCCTCTCTGATGAGCTGTTGT 615
Qy 626 ATGTCAGAACTAAGTCAAGAAATGATCTTATAGAGAGGTTAAATAATGATATGAT 685
Db 616 ATGTCAGAACTAAGTCAAGAAATGATCTTATAGAGAGGTTAAATAATGATATGAT 675
Qy 686 CTTTATTTTGAATTTTGTTCCTCAATATTTGATGATGAGAGTGTGATCTACAGT 745
Db 676 CTTTATTTTGAATTTTGTTCCTCAATATTTGATGATGAGAGTGTGATCTACAGT 735
Qy 746 GAGTTCTAGAGAGCCCACTAGCTATCTGAGACAAATGSCAAAGCTGACATTTGGCT 805
Db 736 GAGTTCTAGAGAGCCCACTAGCTATTTAGAGCAATGSGGAAAGCTGAAATGTGCTC 795
Qy 806 ATTGGAAGTACTGAGATTTTCAATTTCTCACCACTCTTACCAATGTTGAGTTGTT 865
Db 796 ATTGGAAGTACTGAGATTTTCAATTTCTCACCACTCTTACCAATGTTGAGTTGTT 855
Qy 866 GAGAGCTCACTGCAACCTGCCAAACCCCTACCGAAGAAATGAGAGTTTGTCCAG 925
Db 856 GAGAGCTCACTGCAACCTGCCAAACCCCTGCTAAGAAATGAGAGTTTGTCCAG 915
Qy 926 AGCTCTGAGAGAAATGTTGTTGTTTCTTCTGAGGTTGATGCTGATTAACACTCA 985
Db 916 AGCTCTGAGAGAAATGTTGTTGTTTCTTCTGAGGTTGATGCTGATTAACACTCA 975
Qy 986 GAAGAAAGGCGCAATGTAATTTGATGAGCTGAGCTTGCACCAAGATCCCAAAAGTTCTG 1045
Db 976 GAAGAAAGTGCACATGATTTGATGAGCTGAGCTTGCACCAAGATCCCAAAAGTTCTG 1035
Qy 1046 AGATTTGATGGAATTAACAGATCTTAGAGCTCAATACTGCGCTGTACAAGTGATA 1105
Db 1036 AGATTTGATGGAATTAACAGATCTTAGAGCTCAATACTGCGCTGTACAAGTGATA 1095

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 05:23:08 ; Search time 2466 Seconds
(without alignments)
16262.119 Million cell updates/sec

Title: US-09-980-729B-10
Perfect score: 1650
Sequence: 1 agcaactcggaacaacagcat.....ccaattcaagaagacctg 1650

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 1215238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	961.8	58.3	1946	11 AK050435	Mus muscu
2	955.4	57.9	2573	11 AK004971	Mus muscu
3	890.8	54.0	2895	11 AK083294	Mus muscu
4	886.8	53.7	1892	11 AK034801	Mus muscu

5	864.2	52.4	1896	11 AK002736	AK002736 Mus muscu
6	785.8	47.6	1783	11 AK050327	AK050327 Mus muscu
7	781	47.3	2099	11 AK008601	AK008601 Mus muscu
8	769	46.6	2575	11 BC048920	BC048920 Mus muscu
9	606.8	36.8	983	13 BX444042	BX444042 BX444042
10	600.6	36.4	960	13 BQ713091	BQ713091 AGENCOURT
11	599.8	36.4	1004	13 BQ942104	BQ942104 AGENCOURT
12	580.8	35.2	595	10 BG562901	BG562901 602581752
13	576.4	34.9	814	10 BG428781	BG428781 602500860
14	575.4	34.9	927	13 BQ925596	BQ925596 AGENCOURT
15	573.4	34.8	588	2 HSM087819	BX494989 Homo sapi
16	560	33.9	560	14 CB153390	CB153390 K-EST0210
17	546.4	33.1	558	14 CB113345	CB113345 K-EST0155
18	545.8	33.1	559	2 HSM087698	BX494988 Homo sapi
19	545.8	33.1	697	10 BG427192	BG427192 602494236
20	543.8	33.0	560	2 HSM087740	BX494910 Homo sapi
21	529.8	32.1	823	10 BF689099	BF689099 602185172
22	527.2	32.0	930	13 BQ934372	BQ934372 AGENCOURT
23	523.6	31.7	977	12 BI311895	BI311895 602982580
24	522.8	31.7	785	10 BG429706	BG429706 602493711
25	515.6	31.2	891	10 BG189133	BG189133 RST8173 A
26	511.8	31.0	662	14 CB148234	CB148234 K-EST0204
27	511.6	31.0	770	10 BG201991	BG201991 RST21337
28	510.8	31.0	748	10 BG431126	BG431126 602498753
29	506.8	30.7	510	2 HSM073698	BX483530 Homo sapi
30	496.4	30.1	1030	13 BQ714919	BQ714919 AGENCOURT
31	492.6	29.9	746	10 BF688309	BF688309 602185172
32	488.2	29.6	787	10 BG430305	BG430305 602502309
33	485.8	29.4	901	10 BG400539	BG400539 602464778
34	484.4	29.4	756	10 BG429801	BG429801 602494817
35	483.8	29.3	728	10 BG433410	BG433410 602501120
36	480.8	29.1	913	13 BQ714368	BQ714368 AGENCOURT
37	477.8	29.0	801	10 BG212458	BG212458 RST32045
38	471.6	28.6	1005	10 BG167171	BG167171 602344696
39	465.8	28.2	618	14 CB113377	CB113377 K-EST0155
40	460.6	27.9	692	14 CB154247	CB154247 K-EST0212
41	458.8	27.8	759	12 BI330877	BI330877 602981265
42	456.8	27.7	821	10 BG402643	BG402643 602465575
43	450.2	27.3	871	13 BQ900474	BQ900474 AGENCOURT
44	443	26.8	923	9 AI529976	AI529976 u187g04.y
45	441.4	26.8	938	13 BQ942682	BQ942682 AGENCOURT

ALIGNMENTS

RESULT 1
AK050435
LOCUS
DEFINITION Mus musculus adult male liver tumor cDNA, RIKEN full-length
ACCESSION AK050435
VERSION AK050435.1 GI:26341183
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE 10349636
PUBMED
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

QY 430 AATAGAACTATGAGAACTACAGAGCTCAAGATTGATGTTCTTGAGAGATGC 489
 Db 432 GAACAAGAGCTCATGAAAAAACTCCAGAGATCTAAGTTGATGCTGTTAGAGAGATGC 491
 QY 490 TGTTCCTCCCTTGTGTGAGCTGCTGCGGAGTTACTTAAATACCCTTGTCTACAGCT 549
 Db 492 CTGTGTTCCCTGTGTGAGCTGCTATCAGAACTGCTTAAGACACCTTTAGTATACAGTCT 551
 QY 550 CCGCTTCTCTGCTGAGCGCAATTGAAAAGCATATGAGAGAGCTGCTCCCTCCCTC 609
 Db 552 CCGCTTCTCTGCTGAGTACAAATGTAAGAGTACAGTGGGAGCTTCCATCTCCCTCCTC 611
 QY 610 CTATGTCCTGTTGTTATGTCAGAACTAGTACCAATGACTTTCATAGAGAGGTAAA 669
 Db 612 CTATGTCCTGTTGTTGTCAGAACTAGTACCAATGACTTTCATAGAGAGGTAAA 671
 QY 670 AATATGATCTATGCTTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGAGTG 729
 Db 672 GAATATGTTGAGAGGTGTTGCTTTTGAATTTTGGTCCAAATATTTAAGAGAAATCCTG 731
 QY 730 GGATCAGTTCTCAGTGAAGTTCTAGAGAGCCAGTACGTTATCTGAGAGAAATGGCAA 789
 Db 732 GAATCAGTTCTCAGTGAAGTTCTAGAGAGCCAGTACCAATTTAAGATGATGGGAAA 791
 QY 790 AGTGAACATATGCTTATCGAAACTACTGGGATTTTCAATTTCTCACCACCTCTTACC 849
 Db 792 GCGAGACATATGCTGCTTCCGACCTTCTGGGACTTGAATTTCTCACCCTTCTTGGCC 851
 QY 850 AATATGATGTTGTTGAGAGACTTCCACTGCAAACTTGGCAAACTCCCTACCGAAGAAAT 909
 Db 852 TAAATTTGACTTTGTTGAGAGACTTCCACTGTAACCAAGCCAAACACGCTTAAGAAAT 911
 QY 910 GGAAGACTTTGTCAGAGCTCTGAGAGAAATGCTGTTGTTGTTTCTCTGGGGTGAT 969
 Db 912 GGAAGAAATTTGTTGAGAGCTCTGAGAGAAATGCTGTTGTTGTTTCTCTGGGGTGAT 971
 QY 970 GGTGAGTACACGTCAGAGAGAGGCAATGTAATTCATCAGCCCTTGGCAAGATCCC 1029
 Db 972 GGTAAAAACATTAAGAGAGAGGCAATGTAATTCATCAGCCCTTGGCAAGATCCC 1031
 QY 1030 ACAAAGGTTCTGTGAGAGATTTGATGGGAATTAACAGATATCTTGAAGACTCAATCTC 1089
 Db 1032 ACAGAGGTTCTGTGAGAGATTTGATGGGAATTAACAGATATCTTGAAGACTCAATCTC 1091
 QY 1090 GCTGTACAGTGTGATCCCGAGATGATCTTCTGTTGTTGTTTCTCCAAACCAAGCTTTTAT 1149
 Db 1092 GCTGTACAGTGTGATCCCGAGATGATCTTCTGTTGTTGTTTCTCCAAACCAAGCTTTTAT 1151
 QY 1150 CACTCATGTGTGAATGAATGGATCTATGAAGCTATTACCATGGGGTCCCTATGGTGG 1209
 Db 1152 AGCTCATGTGTGAATGAATGGATCTATGAAGCTATTACCATGGGGTCCCTATGGTGG 1211
 QY 1210 AGTTCCATATTTGTTGATGATCAGTTGATACAGTCACTGAAGGCAAGAGAGAGC 1269
 Db 1212 TATTCCTGTTGTTGGGATCAACCTGATTAATTAACACATAGAGCAAGGAGAGAGC 1271
 QY 1270 TGTGAAATTAATCTCAAACTATGACAGAGGAAATTTACTGAGGGCTTTGAGAACT 1329
 Db 1272 TGTGAGTGTGAGCTTGTATCAATGTCAATACAGAGACCTTCTCAGCTGCTTGAAGAGCTGT 1331
 QY 1330 CATTAACGATTCCTCTTAAGAGAAATGATGATGATTAACAGAAATTCACCATGATCA 1389
 Db 1332 CATTAACGATTCCTCTTAAGAGAAATGATGATGATTAACAGAAATTCACCATGATCA 1391
 QY 1390 ACCTGTAAGGCTCTAGATCAGAGCTTCTGATGAGTGTGATGAGTGTGATGAGTGTG 1449
 Db 1392 GCGCATGAGGCTTGTGAGCGAGGCTTCTGATGAGTGTGATGAGTGTGATGAGTGTG 1451
 QY 1450 AGCCAAAGCACTGGATGATGCTGCCCATGACCTCACTGTTCCAGCACTAATCTTAAGA 1509
 Db 1452 AGCCAAAGCACTGGATGATGCTGCCCATGACCTCACTGTTCCAGCACTAATCTTAAGA 1511
 QY 1510 TGTGATTTGGGTTCCGTGAGACCTGTGTGGCAACTGCTATATCTTGTTCACAAATGTTT 1569

Db 1512 TGTGATTTGGGTTCTCTATTTGCTGTGTGAGTGTGATTCATCTCAATGCAAAATGTTG 1571
 QY 1570 TTTATTTCTCTGCAAAAATTTAATAAATACTAGAAATATGAAAA 1614
 Db 1572 CCTCTTTGTTGGCATTAAGACTGCTAACATGAGGAAAAA 1616
 RESULT 2
 AK004971
 LOCUS
 DEFINITION
 Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:130012D20 product:UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSMAL (EC 2.4.1.17) (UDPGr) (UDPGr-2) homolog [status: nonregions], full insert sequence.
 ACCESSION
 AK004971
 VERSION
 AK004971.1 GI:12836559
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kishinai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsumura, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Peele, G., Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Guelinckh, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombetta, P., Nordone, P., Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, L., Wyszynski, B., A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohzuki, S. and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 MEDLINE
 PUBMED

REFERENCE
AUTHORS 5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
AUTHORS
6 (bases 1 to 2573)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hitoaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kaukawa, T., Kato, H., Kawai, J., Kojima, Y., Komori, H., Konda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@ec.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'] GAGAGAGAGAGCGCGCACTGAGTGTGTTTGTGTTTGA 3'), cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5']
GAGAGAGAGAGTCCCAAGCTCATTAATTAATTAACCCCCCCC 3' cDNA was
cleaved with XhoI and SmaI. Cloning sites, 5' end: SmaI; 3' end:
XhoI. Host: SOLR.
FEATURES
source
Location/Qualifiers
1..2573
/organism="Mus musculus"
/mol_type="cDNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:1300012D20"
/db_xref="MGI:1904432"
/db_xref="taxon:10090"
/clone="1300012D20"
/sex="male"
/issue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature
1..2573
/note="UDP-GALACTOSE 4-EPIMERASE 2B1 PRECURSOR,
MICROSOMAL (EC 2.4.1.17) (UDPGR) (UDPGR-2) homolog
[Rattus norvegicus] (SWISSPROT P09875, evidence: FASTY,
85.8%ID, 100%length, match=1587)"
/db_xref="MGI:1919023"
BASE COUNT 733 a 497 g 802 t
ORIGIN
Query Match 57.9%; Score 955.4; DB 11; Length 2573;
Best Local Similarity 75.1%; Pred. No. 7.3e-1591;
Matches 1206; Conservative 0; Mismatches 396; Indels 3; Gaps 1;
QY 13 ACAACATTGATGATAGATGCTATGAAATGAGTTCAGTCTCTCTCTATCA 72
DB 11 AAAAAATTCTCTTCAGCAAGATGCTATGAAACAGGCTTCAGTTTCTTTATTA 70
QY 73 GCTGAGCTTACTTACCTGCGAGTGTGAAAGGTGCTGTGCGCCACAGATT 132

DB 71 GTTCATATGCTATATTAGACCTGAGCCTGTGGAAAAGTGTGTGTGCGCTACAGATA 130
QY 133 CAGCCACGTGATGAATATTAAGAACATCTGATGAACCTGTCAGAGAGGTCAAGAGT 192
DB 131 CAGCATTTGATTAATATGAATAATATCTGATGAACCTGTCAGAGAGGTCAAGAGT 190
QY 193 GACTGTATGACATCTTCAGCTTCATTTCTTGATCCCAAGCCCATCTACTTAA 252
DB 191 CACGTTCTCATATCTTCTGCTTCATCTCATATGGGCCGAGCAATGATCTTATTA 250
QY 253 ATTGAAGTTATCTCTGATCTTTAACTAAACGATTTGAGATATATCAAGAGCT 312
DB 251 TTTGAAATTTATCTCACCCTTGATGAATAAGATCTGAATATCTTTGAAAAATG 310
QY 313 GGTAAAGATGAGGCA---GAACCTCCAAAAGACATTTTGATATTTTTCACAGT 369
DB 311 GGTGGAATCTGACATACGATTAATTAATAAACTTCATTTGACATCTTATTA 370
QY 370 ACAAGAAATCATGTGACATTTAATGAATATTAAGATTTCTGTAAAGATATAGTTTC 429
DB 371 GCAAAAATCTCCAGTGAATATTCAGATGATGAAGTTCTGCAAGCAGTATGTTG 430
QY 430 AATATGAATCTTATGAAGAACTACAGAGTCAAGATTTGATTTGTTCTGAGATGC 489
DB 431 GAAAGAGGCTCATGAAAAAATCCAAAGATCTAAGTTGATGTGTTCTGAGAGATGC 490
QY 490 TGTTCCTCCCTTGTGAGTGTGCGAGTGTACTTAAATACCTTTGCTACAGCT 549
DB 491 CTGCTTCCTGCTGTGAGTGTGCTGTATAGACCTCTTTGATATACAGCT 550
QY 550 CCGCTTCCTTCCTGCTACGCAATTTGAAGAAAGATGAGAGCTTCTGTCCTTC 609
DB 551 CCGCTTCCTGCTGCGGATATCTAATGAAAGTACAGTGGGGCTTCCACTCCCTTC 610
QY 610 CTATGCTCTGTTTATATGTAAGTACATGACCAATATGCTTATCATAGAGGTTAA 669
DB 611 CCAATGCTGCTGCTTCTGTCAGAACTAAGTACATGACATTTGCAAGAAAGTTAA 670
QY 670 AATATGATCATGATGCTTATTTGAATTTGTTGCAATATTTGACATGAAGAGTG 729
DB 671 GAATATGTCAGAGTGTGCTTTTGAATTTGTTGTTCAATTTAAGAAATCTTG 730
QY 730 GGATCAGTTCTACAGTGAAGTTCTAGAGAAAGCCACTATCTGAGCAATGGCAA 789
DB 731 GAATCAGTTTACAGTATGTTCTAGGAGACCTAACAATTAAGTATGATGAGGAA 790
QY 790 AGCTGACATATGCTTATTCGAAACTACTGAGATTTTCAATTTCTCACCCTTACC 849
DB 791 GGCAGACATATGCTGCTGTCGAACTTCTGAGCTTGAATTTCTCACCCTTTCTGCGC 850
QY 850 AATGTTGAGTTCGTTGAGAGACTCCACTGCAACCTGCCAAACCCCTACGAGGAAT 909
DB 851 TAAATTTGACTTTGTTGAGAGCTCCTATTTGAACCAAGCCAAACCACTGCTTAAGAAAT 910
QY 910 GGAAGAGTTGTCAGAGCTCTGAGAGAAAGTGTGTGTGTTCTCTGAGGATGAT 969
DB 911 GGAAGAAATTTTTCAGAGCTCTGAGAGAAAGTGTGTGTGTTCTCTGAGGATGAT 970
QY 970 GGTGATGAACGCTCAGAGAAAGGCGCAATGATATGATGATGATGATGATGATGAT 1029
DB 971 GGTATTAATAATTAAGAAAGGCGCAATGATGATGATGATGATGATGATGATGAT 1030
QY 1030 ACAAAAGTTGCTGAGAGATTTGATGGAATTAACCAATCTTTAGGACTCAATATCG 1089
DB 1031 ACAGAGGTTGCTGAGAGATTTGATGGAATTAACCAATCTTTAGGACTCAATATCG 1090
QY 1090 GCTGTAAGTGAATATCCCAAGATGATCTTTGATGATGATGATGATGATGATGAT 1149
DB 1091 GCTCTACAGTGAATCCCAAGATGATCTTTGATGATGATGATGATGATGATGAT 1150
QY 1150 CATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209
DB 1151 AGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1210

QY 1210 AGTCCCATATTTGGTATGATGCTTGTATACATAGCTACATGAAGGCCAAGAGCAGC 1269
 DB 1211 TATTCCTCTGTTGGGATGATCACTGATATATTAACCACTATAGCCAGAGACAGC 1270
 QY 1270 TGTAGAAATTAACCTTCAAACTATGACAGGAAAGATTACTGAGGGCTTTGAGACAGT 1329
 DB 1271 TGTAGAGTGAAGCTTATGATCAATGATCAATACAGACCTTCTCATGCTGTAAGACTGT 1330
 QY 1330 CATTAACGATTCCTCTTATATAGAAATGCTATGATATATCAAGATTCACCATGATCA 1389
 DB 1331 CATTAATGACCTTCCTCTTATATAGAAAGAACGCCATAGATTAACCAATCCATGACCA 1390
 QY 1390 ACTGTAAAGCCCTAGATGAGAGCTTCTGTGATGAGTGTGTCATGAGCCCAAGG 1449
 DB 1391 GCCAATGAAGCCCTTGAACGACGCTGCTTGTGATGAGTATGTCTATGACCAAGAGG 1450
 QY 1450 AGCCAGACCTCTGATGATGAGTGGCCATGACCTGATGAGCTTCTATGAGCACTATAGA 1509
 DB 1451 AGCCAGACCTCTGATGAGCTGGCCATGATGACCTTACCTGATGAGCACTCTGGA 1510
 QY 1510 TGTGATTTGGTCTCTGATGAGCTGTGTGAGCACTGCTATATCTTTGACAAATGTTT 1569
 DB 1511 TGTGATTTGGTCTCTGATGAGCTGTGTGAGCACTGCTATATCTTTGACAAATGTTT 1570
 QY 1570 TTTATTTTCTGTCAAAATTTATTAATAAGTAAAGTAGAGAA 1614
 DB 1571 CCTTTTGTGTCATAGACTGCTAACATGAGGAAAGAGAGAA 1615

RESULT 3
 AK083294
 LOCUS
 DEFINITION
 Mus musculus adult male liver tumor cDNA, RIKEN full-length
 enriched library, clone: C730031G17 product: similar to
 UDP-GLUCURONOSYLTRANSFERASE 2B5 PRECURSOR, MICROSMAL (EC 2.4.1.17)
 (UDPgt) (M-1) [Mus musculus], full insert sequence.
 AK083294
 AK083294.1 GI:26350414
 HTG: CAP trapper.
 HTG: CAP trapper.
 Mus musculus (house mouse)
 Mus musculus
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

REFERENCE
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Glasi, C., King, B., Kochiwa, H.,
 Kiehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schiraldi, L. M., Staudt, F., Suzuki, R., Tomita, M.,
 Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bona, M., Brownstein, M. J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilting, L.,
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohseki, S.
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 11217851

REFERENCE
 5
 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2895)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

COMMENT
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

FEATURES
 source
 1. 2895
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:C730031G17"
 /db_xref="taxon:10090"
 /clone="C730031G17"
 /sex="male"
 /cell_type="tumor"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"

URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>
 location/Qualifiers

Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissue was provided by William A. Held, Roswell Park Cancer
 Institute, Department of Molecular and Cellular Biology, Elm and
 Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>

CDS

```

/dev stage="adult"
33..1632
/feature="untranslated protein product: putative
similar to UDP-GLUCURONOSYLTRANSFERASE 2B5 PRECURSOR,
MICROSOMAL (EC 2.4.1.17) (UDPgt) (M-1) [mus musculus]
(SwissProt|P17171, evidence: FASTA, 82.6%id, 100%length,
match=1587)"
/codon_start=1
/protein_id="BAC38847.1"
/db_xref="GI:26350415"
/translation="MPVKMISALLLLQNSCFRSTSGKLVLPFESHMNLKILID
ELVORGEVIVLRPSASIFVDPKYSFGIKETPTASKDYLETFLKLVDEMFEV
RDTCYSYLQTLFGLSDYSLSCKEASVKOMATKQESKFLVILDMASGCEIL
AOLIOPLVLSRSPGVQVEKNSGFLPSYVPLVLSGGQMTFERKIMCML
VFDMFQTFTEKEDQPSFETIGRPTLITMGKAEWAFISVYDLEPHTLPLNVEY
VGGHCKRPFKEMEDFVOSGHHGVVSLSNMNSMTBERANISALAIPOK
VLMRFDGTPASLGNTRIVYMLPONDILGPKTKAPLTHGANGLEYAHLGIPMG
IPLPSEHDNIHNVAKTAVANLIRMSRDLNLALEVINNSYKENVMLSTIHH
DQMKPLNRTIFMIEFVWRHKGAHLRPLAHLNLTWYQVHSLDVLGVATVAVLP
VKCLFVYGFVKAKKKKXNE"
BASE COUNT      824 a      566 c      607 g      898 t
ORIGIN
Query Match      54.0%; Score 890.8; DB 11; Length 2895;
Best Local Similarity 72.3%; Pred. No. 3e-177;
Matches 1156; Conservative 0; Mismatches 442; Indels 0; Gaps 0;

```

```

31 CAGATGCTCTAATGAATGACCTTCTCTCTGATACAGCTGACTGATTTAG 90
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
29 CAGATGCTCTGAAAGTGAATTCCTCTCTCTGATACAGCTGACTGATTTAG 88
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
91 CTCTGGAGTGTGAAAGTGTCTGTGGCCACAGAAATTCAGCACTGATGATAT 150
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
89 ATCCACCGCTGTGGAGAGTGTGTGTGGCCACAGAAATTCAGCTGATGATAT 148
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
151 AAAGCAATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 210
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
149 AAAGTATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 208
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
211 AGCTTCCATTTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 270
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
209 AGCTTCCATTTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 268
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
271 ATCTTCACTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 330
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
269 AGCTTCACTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 328
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
331 ACTTCCAAAGACATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 390
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
329 TGAAGTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 388
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
391 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
389 ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 448
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
451 ACTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 510
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
449 ATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
511 GCTGCGCAGTACTTAAATGATGATGATGATGATGATGATGATGATGATGAT 570
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
509 CATAGCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
571 AATGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 630
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
569 AGTAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
631 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 690
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
629 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 688
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
691 TTTGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 689 TTTGATCTTTGTTCCAGATTTTACAGAAAGAAATGGATCAATTTTACAGTGAAC 748
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 751 TCTAGAAAGACCACTACATGTTATCTGAGCAATGCGAAAGATGATGATGATGAT 810
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 749 TTTGGAGAGGCCCAACCTTAAATGAGCAATGGGAAAGCAAAATGATGATGATGAT 808
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 811 AAATCTGAGATTTTCAATTTTCTCCAGCACTTTACCAATGATGATGATGATGATGAT 870
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 809 ATCTCACTGGGATCTGAGATTTTCTTCCACCACTTACCAAAATGATGATGATGATGAT 868
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 871 ACTTCATGCAAACTTCCCAAAACCCCTTCCAGAAAGAAATGAAAGATTTTCCAGACTC 930
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 869 ACTTCATGCAAACTTCCCTAAGCCCTTCCAGAAATGAAAGATTTTCCAGACTC 928
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 931 TGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 990
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 929 TGGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 988
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 991 AAGGCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1050
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 989 AAGGCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1048
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1051 TGAATGCAATTAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1110
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1049 TGAATGCAATTAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1108
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1111 GAATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1170
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1109 GAATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1168
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1171 GATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1230
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1169 ACTCTAGAGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1228
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1231 GCTTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1290
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1229 ACATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1288
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1291 TATGACAAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1350
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1289 AATGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1348
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1351 AGAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1410
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1349 AGAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1408
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1411 AGCATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1470
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1409 GACTATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1468
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1471 TGCCATGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1530
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1469 TGCCATGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1528
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1531 CTGTGTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1590
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1529 CTGTGTGTCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1588
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1591 TAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1628
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1589 TGTAAAGAAAGCAAAAGAAAGAAATGATGATGATGATGATGATGATGATGATGAT 1626
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 4
 AK034801 1892 bp mRNA linear HTC 05-DEC-2002
 LOCUS
 DEFINITION
 Mus musculus 12 days embryo body between diaphragm region
 and neck cDNA, RIKEN full-length enriched library, clone:9430041C03
 product:UDP-glucuronosyltransferase 2 family, member 5, full insert
 sequence.
 ACCESSION
 AK034801
 VERSION
 AK034801.1 GI:26330209
 KEYWORDS
 HTC; CAP trapper.

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	2 10349636
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	4 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojoubori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, L., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldridge, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, V., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 1892)
PUBMED	12101892
REFERENCE	6
AUTHORS	6 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kuibara, C., Matsunura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numaizaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

TITLE	Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
COMMENT	Submitted (16-JUL-2001) Yoshitake Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
FEATURES	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details.
source	URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location: Qualifiers
source	1. 1892 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:9430041C03" /db_xref="taxon:10090" /clone="9430041C03" /tissue_type="embryonic body between diaphragm region and neck" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 34. 1626 /note="unnamed protein product; UDP-glucuronosyltransferase 2 family, member 5 (MGI:98900) putative" /codon_start=1 /protein_id="BAC28835.1" /db_xref="GI:26330210" /translation="MPGKMSIALLLQISCCPSRYKGVLPMPFESHMNIKITLD ELVORERTVTPRAAPVYVILDPKRSGLGKPEPVSVDNLENPIKRWDTVMP RDTCLSYSLQWMDIBSDYFSLCKDVNSKELMTKQESFVLDLDPVASCBL LAEILLDPLPLYSIRFSPQYIEKSSGFLLPSPVYVILSLGGQMTFERVNMCR LVEFDWFQNFNRKMDSPFSEYLGRTPLTAEETKXEMMLIRSNMDLEFPPLPVD YGLGKLCRAKPLPKDMEFVOSGDGVVVFSLGSVNMTEKNTAMALAOIPIQ KYLWKGDKTPPLTIGNTRVVKMLPONDILGHFKTAPVTHGGANGVEVYIYIGIMI GIPLEBQNDNTAHVAKGAATVNTIRMSRSDVNALSEVIDNPEYKNAITLSLTH HDDPIPLRAVFWFVFRHKRAKRLSLGHLTVHQYHFLDVLGFLSCAVTVLV TVKCLPLIFRFPVKEKKIKNE"
BASE COUNT	563 a 356 c 393 g 580 t
ORIGIN	
Query Match	53.7%; Score 886.8; DB 11; Length 1892;
Best Local Similarity	72.5%; Pred. No. 2.1e-176;
Matches 1162; Conservative	0; Mismatches 437; Indels 3; Gaps 1;
QY	30 TCAGATGCTAAGTAAGTGAAGTCTTCTTCTGATACAGCTGAGCTGTAATCTTA 89
DB	29 TCGATGCTGGAAGTGAAGTGAAGTCTTCTTCTGCTGCTGCGATGATGATGCTTCA 88
QY	90 GCTCTGGAAGTGTGGAAGTGTGCTGTGTGCGCCACAGAAATTCAGCCATGATGA 149
DB	89 GATCTGGAAGTGTGGAAGTGTGTGTGTGTGCGCATGGAATTCAGTCATGATGACA 148
QY	150 TAAAGCAATCTGGAAGTGAAGTGTGTCAGAGAGAGTCTAGAGTGATGATTTGGCAT 209
DB	149 TAAAGTAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 208
QY	210 CAGCTTCATTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 269
DB	209 CAGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 268
QY	270 TATCTTAACTAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 327
DB	269 CATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 328

Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

21085660

11217851

5

THE FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 1896)

Adechi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kaubawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Further visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAGAGCGCGCGCACTCGAGTGTGTTTTTTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to R₀C = 5.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGAGAGGATCCAGAGCTCAATTAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

Source

Location/Qualifiers

1..1896

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM,DB:0610033E06"

/db_xref="MGI:1904280"

/db_xref="taxon:10090"

/clone="0610033E06"

/sex="male"

/tissue_type="Kidney"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

36..1629

/note="UDP-glucuronosyltransferase 2 family, member 5 (MGI:MGI:98900)

putative"

/db_xref="MGI:98900"

/db_xref="MGI:1904280"

misc_feature

571 a _361 c 386 g 578 t

BASE COUNT

571 a 361 c 386 g 578 t

ORIGIN

Query Match

Best Local Similarity

52.4%; Score 864.2; DB 11; Length 1896;

71.7%; Pred. No. 1,2e-17;

Matches 1162; Conservative 0; Mismatches 453; Indels 6; Gaps 2;

QY TCAGAGTGTCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGAGCTGTACTTGA 89

30 TCAGAGTGTCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGAGCTGTACTTGA 89

31 TCAGAGTGTCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGAGCTGTACTTGA 90

QY GCTCTGGAAGTTGTGAAAGTGTGTGTGTGAGCCACAGAAATTCAGCACTGATGATA 149

90 GCTCTGGAAGTTGTGAAAGTGTGTGTGTGAGCCACAGAAATTCAGCACTGATGATA 149

91 GATCTGTGAATGTGGAAAGTGTGTGTGTGAGCCACAGAAATTCAGCACTGATGATA 150

QY TAAAGACATCTCGATGAACTGTGTCAGAGAGCTATGAGGACTGTATGGACATTT 209

150 TAAAGACATCTCGATGAACTGTGTCAGAGAGCTATGAGGACTGTATGGACATTT 209

151 TAAAGACATCTCGATGAACTGTGTCAGAGAGCTATGAGGACTGTATGGACATTT 210

QY CAGCTTCATTTCTTTGATCCACAGAGCCATCTAATTTGAATTTGAATCTCTG 269

210 CAGCTTCATTTCTTTGATCCACAGAGCCATCTAATTTGAATTTGAATCTCTG 269

211 CAGCTTCATTTCTTTGATCCACAGAGCCATCTAATTTGAATTTGAATCTCTG 270

QY TATCTTTAATCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGATGAGCT 327

270 TATCTTTAATCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGATGAGCT 327

271 CATCTGCACTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGATGAGCT 330

QY -AGAACTTCAAGACATTTTGTGATATTTTCAAGATGACAAAGAAATGATGGA 386

328 -AGAACTTCAAGACATTTTGTGATATTTTCAAGATGACAAAGAAATGATGGA 386

331 ATGAGTTGTCAAGAGATACATGTTTGTATTTCTCTTAATGCAAAATATGTTGATG 390

QY CATTTATGACATCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGATGAG 446

387 CATTTATGACATCTAAGAAATGAGACTTCAGCTCTCTCCGATACAGCTGATGAG 446

391 AATATCTGTATTTATCTAAGCTTTGTGAAGTGTGTTCAAGACAGCTGATGAG 450

QY AGAACTTCAAGACATTTTGTGATATTTTCAAGATGACAAAGAAATGATGGA 506

447 AGAACTTCAAGACATTTTGTGATATTTTCAAGATGACAAAGAAATGATGGA 506

451 CAAGCTTCAAGACATTTTGTGATATTTTCAAGATGACAAAGAAATGATGGA 510

QY AGCTGTGCGCGAGTACTTAATTAATCCCTTGTCTACAGCTCTCTCTCTGCT 566

507 AGCTGTGCGCGAGTACTTAATTAATCCCTTGTCTACAGCTCTCTCTCTGCT 566

511 AGCTAATCGCTAATTTCTCAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 570

QY ACGAATTTGAAGAAATGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 626

567 ACGAATTTGAAGAAATGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 626

571 ACGAATTTGAAGAAATGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 630

QY TGTGAGAACTAAGTGAACCAATGACTTTCAATAGAGAGGTTAAATATGATCTATG 686

627 TGTGAGAACTAAGTGAACCAATGACTTTCAATAGAGAGGTTAAATATGATCTATG 686

631 TGTGAGAACTAAGTGAACCAATGACTTTCAATAGAGAGGTTAAATATGATCTATG 690

QY TTTATTTGAATTTGTGTTCCAAATATTTGATGAGAAAGTGTGATCTTCACTG 746

687 TTTATTTGAATTTGTGTTCCAAATATTTGATGAGAAAGTGTGATCTTCACTG 746

691 TTTATTTGAATTTGTGTTCCAAATATTTGATGAGAAAGTGTGATCTTCACTG 750

QY AAGTCTAGAGAACCCACTACATCTATCTGAGCAATGCGAAAGTGCATATGCTTGA 806

747 AAGTCTAGAGAACCCACTACATCTATCTGAGCAATGCGAAAGTGCATATGCTTGA 806

751 AATATTTGGAAGAGCCCAACCACTTACAGCAAGTGTGAGTGTGCTTGA 810

QY TTGCAAACTAAGTGAATTTTCAATTTCTGACCACTTCAACCAATGTTGATGCTTGA 866

807 TTGCAAACTAAGTGAATTTTCAATTTCTGACCACTTCAACCAATGTTGATGCTTGA 866

811 TTGATCTCAACTGGAATTTGAGATTTCTTCTGACCACTTCAACCAATGTTGATGCTTGA 870

QY GAGAGCTCACTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTG 926

867 GAGAGCTCACTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTG 926

871 GAGAGCTCACTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTGCAAACTGTG 930

QY GCTCTGGAAGAAATGAGT 986

927 GCTCTGGAAGAAATGAGT 986

931 GCTCTGGAAGAAATGAGT 990

QY AAGAAAGGCGCAATGTATGATCTAGCTGAGCTTGTGCAAGATCCCAAAAGTGTGTGGA 1046

987 AAGAAAGGCGCAATGTATGATCTAGCTGAGCTTGTGCAAGATCCCAAAAGTGTGTGGA 1046

991 AAGAAAGGCGCAATGTATGATCTAGCTGAGCTTGTGCAAGATCCCAAAAGTGTGTGGA 1050

QY GATTGATGGAATTAACAGATCTTTAGAGCTCAATCTCGGCTGTACAGTGTATAC 1106

1047 GATTGATGGAATTAACAGATCTTTAGAGCTCAATCTCGGCTGTACAGTGTATAC 1106

1051 AATTGATGGAATTAACAGATCTTTAGAGCTCAATCTCGGCTGTACAGTGTATAC 1110


```

FEATURES
    source
        Location/Qualifiers
            1..1783
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="PANTOM_DB:C730037110"
                /db_xref="taxon:10090"
                /clone="C730037110"
                /sex="male"
                /cell_type="tumor"
                /issue_type="liver"
                /clone_id="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="adult"
                /cdev_stage="adult"
            34..1638
                /note="unnamed protein product; putative
                similar to UDP GLUCURONOSYLTRANSFERASE UGT2A3 [Cavia
                porcellus] (SPTRIQ9R10, evidence: FASTA, 71.3%id,
                99.8%length, match=1579)"
                /protein_id="PAC34191.1"
                /db_xref="GI:26341058"
                /translation="MVSEKCAAFPELLQLCAGGSPSKVLVWPCDMSHNLKLTILE
                ELGARHEVTLKTPSIIIDQSKRPLHFNENPLLYEIEAENKLNLIANLVAVIPN
                LSLMEAKTLQDFLOVTDGESICRSVLNOKFMDLRDQYDVVVIDVPCGELV
                AELVQLPEVYTLPEFSGYMEKHGGLPIPLSYVPMSELDTMTETTERKMMESL
                LPEVWLQOYDEAFMDQFSETLGRPTTECTVKGADIMLI RTYMDVEFPPVLPNPEF
                VGGIHCBAKPLKEMEPEFVSGEHWVPSIGSMKKNLTREKANIASVLAQIPQK
                VUNRYSKRPATLGSNTRLNWIRPQNDLHLPKTKAITTGTTGTYEALHYGVPMVG
                DPMIGDQPHNIAHMEAKGAALKVISITMTSIDLSAVRAVINEPSYENMRSLRIH
                DQPKPLRAVFWIEFVWRHKGAHLVAHADLSWFOYHSLDVIGFLLCVLLTFPII
                TKFCLFVCKLYMKESKMGNRKKKN"
                1756..1761
                    /note="putative"
                    1783
                        /note="putative"
                        365 c 416 g 492 t
BASE COUNT 510 a 365 c 416 g 492 t
ORIGIN
    polyA_signal
        1783
            /note="putative"
            1783
                /note="putative"
    polyA_site
        1783
            /note="putative"
            1783
                /note="putative"
Query Match 47.6%; Score 785.8; DB 11; Length 1783;
Best Local Similarity 69.9%; Pred. No. 4e-155;
Matches 1119; Conservative 0; Mismatches 472; Indels 10; Gaps 4;

```

```

    QY 505 TGAGCTGCTGCGGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTCTCTG 564
    DB 504 AGAGTTGTGCGAAGAGTCTTACAGTCCCTTTCGTATACACTGAGTTACAGATGGG 563
    QY 565 CTACGCAATTGAAAGCATAGTGAAGACTTGTTCCTCTCTATGTGCTGTGTG 624
    DB 564 CTACTACATGAGAAACAGTGTGCGCCTTCCAAATTCACCTCGATGACCGGTGTG 623
    QY 625 TAGTCAAACTAAGTGAACCAATGACTTTCTATAGAGAGGTAAATAATGATCTATGT 684
    DB 624 CATGAGTGAAGCTACAGACAAATATGACCTTCAAGAGAGGTGAAATAATATGATTTTC 683
    QY 685 GCTTATTTTGAATTTTGTTCCTCAAAATATTTGACATGAAGAAGGGGATCGATTCTACAG 744
    DB 684 ACTGTTGTTTGAATCTGCTCTCAGCANTATGACTTTGCACTTGGGATCAGTTTTCAG 743
    QY 745 TGAAGTTTGAAGAACCCACTACGTTATCTGAGACATGACAAAGCTGACATATGGCT 804
    DB 744 TGAACCTTGAAGAGGCCCAACGTTCTGTAAAGCTGTGGGAAAGCTGACATTTGGCT 803
    QY 805 TATTCGAATCTAGGATTTTCAATTTCTCACCCACTTTACCAATGTTGAGTTGCT 864
    DB 804 AATCCGAACATATTTGGATGTTGATTTCTCTGTCATATTTACCAATTTTGAATTTGT 863
    QY 865 TGAGGAGCTCACGCAAACTGCGCAACCCCTAACGGAAGAAATGGAAGATTTGTCCA 924
    DB 864 GGAAGACCTCACCTGCAAACTGCGCAACCCCTTTTACCTTAAGAAATGGAATTTGTTCA 923
    QY 925 GAGCTCTGAGAAATATGTTGTGTGTGTTTCTCTGAGGTCGATGTCATACAGTGC 984
    DB 924 GAGCTCTGAGAAATATGTTGTGTGTGTTTCTCTGAGGTCGATGTCATACAGTGC 983
    QY 985 AGAAGAAAGGCCAATGTAATTTGATCAGCCCTTGCCAGATCCCAAAAGTTCTGTG 1044
    DB 984 AGAAGAAAGGCCAATCTCAATGCTCTGTCTCTGTCACCAATTTCCAGAAAGTTGTG 1043
    QY 1045 GAGATTTGATGGGAATTAACCAATGATCTTAGAGCTCAATACCTGGGTGTAAGTGGAT 1104
    DB 1044 GAGATTTGATGGGAATTAACCAATGATCTTAGAGCTCAATACCTGGGTGTAAGTGGAT 1103
    QY 1105 ACCCCGAATGATCTTCTGATGATCCCAAAACCAAAAGCTTTTATCATCATGATGAT 1164
    DB 1104 TCCCCGAATGATCTTCTGATGATCCCAAAACCAAAAGCTTTTATCATCATGATGAT 1163
    QY 1165 GAAATGATCTATGAAGCTATTTTACATGAGGGTCCATATGTTGAGATTTTGG 1224
    DB 1164 AATCGGATTTATGAAGCTATTTTACATGAGGGTCCATATGTTGAGATTTTGG 1223
    QY 1225 TGATCGCTTGAATACATGCTCAATGAGGCGCAAGAGAGAGCTGTAAATTAATCTT 1284
    DB 1224 GATACGCTTGAATACATGCTCAATGAGGCGCAAGAGAGAGCTGTAAATTAATCTT 1283
    QY 1285 CAAATCTATGACAGGAGATTTTACTGAGGGCTTTGAGAGACGATCTTACCGATTCTC 1344
    DB 1284 CAGTACATGACAGGAGATTTTACTGAGGGCTTTGAGAGACGATCTTACCGATTCTC 1343
    QY 1345 TTATTAAGAGATGCTATGAGATTTATCAGAAATTCACATGATCAACTGTAAAGCCCT 1404
    DB 1344 TTATTAAGAGATGCTATGAGATTTATCAGAAATTCACATGATCAACTGTAAAGCCCT 1403
    QY 1405 AGATCGAGAGCTCTTCTGATGAGATTTGATGAGGCGCAAGAGAGAGAGAGAGAGAG 1464
    DB 1404 AGATCGAGAGCTCTTCTGATGAGATTTGATGAGGCGCAAGAGAGAGAGAGAGAGAG 1463
    QY 1465 ATCAGCTGCCATGACCTCACTGTTCCAGACTACTCTATGATGATGATGATGATGATGAT 1524
    DB 1464 TGTGAGAGGCCAATGACCTCACTGTTCCAGACTACTCTATGATGATGATGATGATGATGAT 1523
    QY 1525 GCTGACCTGTGAGCACTGCTATATTTCTTGTTCACAAATGTTTTTATTTTCTGTCA 1584
    DB 1524 ATTGTTGTGTGTGTTTCTGATGATCACTCAATTAATTTGTTGTTGTGTGTCA 1583

```

Oy	1585	AAATTTAATAACTAGAAAATGATGAAGAAGCAATAGA	1625
Dd	1584	AAAACTTATTATGCA-MAGAAAAGTAAGAAAATCGGGACACGA	1623
RESULT 7			
AK008601			
LOCUS			
DEFINITION	AK008601	2099 bp	mRNA linear HTC 05-DEC-2002
	Mus musculus adult male small intestine cDNA, RIKEN full-length		
	enriched library, clone:2010321U07 product:similar to UDP		
	GLUCURONOSYLTRANSFERASE UGT2A3 [Cavia porcellus], full insert		
	sequence.		
ACCESSION	AK008601		
VERSION	AK008601.1	GI:12842684	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
TITLE	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS			
TITLE	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
JOURNAL	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to		
PUBMED	prepare full-length cDNA libraries for rapid discovery of new genes		
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)		
AUTHORS			
TITLE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,		
JOURNAL	Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Teshito, H., Itoh, M.,		
MEDLINE	Sunji, N., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A.,		
PUBMED	Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K.,		
REFERENCE	Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,		
AUTHORS	Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,		
TITLE	Yonekita, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format		
MEDLINE	sequencing pipeline with 384 multicapillary sequencer		
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)		
REFERENCE	20530913		
AUTHORS	11076861		
TITLE	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,		
JOURNAL	Arikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,		
MEDLINE	Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, T.,		
PUBMED	Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R.,		
REFERENCE	Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,		
AUTHORS	Fleischmann, M., Gaasterland, T., Glass, C., King, B., Kochiya, H.,		
TITLE	Kuvel, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,		
JOURNAL	Kuekenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,		
MEDLINE	Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,		
PUBMED	Baldarelli, R., Barsh, G., Blake, J., Botfield, D., Bojunga, N.,		
REFERENCE	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bul, C.,		
AUTHORS	Pfecher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,		
TITLE	Hofman, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,		
JOURNAL	Marchionni, L., Mashima, J., Mazzarelli, J., Momberte, P., Nordone, P.,		
MEDLINE	Ring, B., Schawalid, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,		
PUBMED	Sato, K., Shimbach, C., Seya, T., Shibata, Y., Storck, K.F., Suzuki, H.,		
REFERENCE	Toyoh-oke, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,		
AUTHORS	Myoshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsuke, S.		
TITLE	and Hayashizaki, Y.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
MEDLINE	Nature 409 (6621), 685-690 (2001)		
PUBMED	21085660		
REFERENCE	11217851		
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research		
TITLE	Group Phase I & II team.		

TITLE
JOURNAL
REFERENCE
AUTHORS

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2099)

Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carlinici,P., Fukuda,S., Fukushima,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirakawa,T., Horii,F., Imotani,J., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Koys,S., Kuwihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numezaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sekai,K., Sano,H., Saeki,D.,
Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-1-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel.:81-45-503-9222,
Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGGAGAAGGAATCCCAAGACCTTTTGTTCCTTTTTCVN 3'], cDNA was
prepared by using trihalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and Subtraction to
Rot = 20.0. Second strand cDNA was prepared with the primer adapte
of sequence[5'
GAGGAGAAGATTCTCGAGTTAATTAATAATGCCCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SacI. Cloning sites, 5' end: XhoI, 3' end: SacI.
Host: SOLR.

location/qualifiers

FEATURES
source

1..2099
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:2010321J07"
/db_xref="MGI:1904826"
/db_xref="taxon:10090"
/clone="2010321J07"
/sex="male"
/tissue-type="Small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
15..1619
/note="unnamed protein product; putative
similar to UDP GLUCURONOSYLTRANSFERASE UGT2A3 [Cavia
porcellus] (SEPR|Q9RL10, evidence: PASTY, 71.3%ID,
99.8%length, match=1579)"
/codon_start=1
/protein_id="BAB25770.1"
/db_xref="MGI:1284286"
/db_xref="MGI:1919344"
/translation="MVSEKCVAAPFLIOLCMAGGFSGSKTVMPDGMHWMLTKLTILHE
ELGRGHETVLDPPLIIIDSKRIPIHFENIPLLILEITENFLNLANLAVNPICGL
LSLEAKATLDLPLOVYGDESI CRSLVNVQKMDKRDQAQVVIVDVPAVCGL
AEVQIPPVYLRSMGYMEKHGCGLPIPSVPVMSELTDMMTEPERYKNMFIS
LEFWILOOYADPDOPFSETIGRPTECKGVGADWLIRTYMDEVPEPYLVPFEEF
VGGHCRKPARKLPREMEEFNSGEHVGSLSMTWKNLTKEANLIASVLAOIPOH
VLMVSGKKPILSNTRLFNMWITPONDLIGHPKTKPTTHGCTNGIYAIAYGVPMN
VPMGGDDPHNTAHMBARGAALKAVISITWTSTDLSAVAVINBSYKFNARLSIH
DQPKPLDRAVFWIEFWNRHGAKLRVAAHLDSWFQYHSLDVIGFLICVTTLTFII
TKFCIFCDKRYMESKKMGRRKKKN"

CDS

618 a 406 C 467 g 608 t

ORIGIN

Query Match 47.3%; Score 781; DB 11; Length 2099;
 Best Local Similarity 69.7%; Pred. No. 4.2e-155;
 Matches 1116; Conservative 0; Mismatches 475; Indels 10; Gaps 4;

28 CATGAGATGCTATGAAATGAGACTTCAGCTCTTCTCCTGATAGAGCTGAGCTGTAATT 87
 11 CAATATGCTCTGAAAAATGTTGGGCAATTTTCTGCTGCAAGCTTTCTGGGCC-- 68
 88 TAGCTTGGGAGTTGTGAAAGGTGCTGTGTGCCCCAGAAATTCAGCCTGAGTAA 147
 69 -GGCTGTGGATTCTGCGCAAGGTCTCTGTGGCCCTGTATATGAGCCACTGGCTGAA 127
 148 TATAAGACATCTCGATGATGAACTTGTCCAGAGGTATGAGGATGATGATGGCATC 207
 128 TCTAAGACTATCTTCTGAGAGCTTGGAGCAAGAGGCGACAGAGTAAAGTCTGAAATA 187
 208 TTGAGCTTCCATTTCTTTCGATCCCAAGCCCACTCTACTCTTAAATTTGAAGTTATCC 267
 188 CCCAGATCAT---CATAGATCAGAGTAAACGTATTCACATGCACTTGTGAGATATTCC 244
 268 TGTATCTTTAACTAAACAGTGTGAGATATTAATCAAGCAGCTGTGTAAGAGA--TG 324
 245 TTGCTGTATGAATTCAGACAGCTGAGAAATGTTTAAATGAGATGCAATCTAGCTGT 304
 325 GCGAGAACTTCCAAAGACATTTTGTGCTATTTTTCACAAGTACAAAGAAATCATGTG 384
 305 GAATGTCAATTCAAACCTGTCTACTGTGGAGAGCAAAAAATTACAAAGCTTCTTCT 364
 385 GACATTTATGACATTAAGTAAAGTCTGATAGATATGATTTCAATAAGAACTTAT 444
 365 TCAAGTAACTGAGATTTTGAAGATTTTGAAGAGTATATTGACACGAAATTCAT 424
 445 GAAAGAACTACAGAGATCAAGATTTGATGTTCTTGCAGATGCTGTTTCCCTTTGG 504
 425 GGAAGAGTACGGAGTGCACAAATATGATGATGATGATGATGATGATGATGATGATG 484
 505 TGAGCTGTGGCCGAGTACTTAAATACCTTTGTCTACAGCTTCCGCTTCTCTCTG 564
 485 AGAGTTGTGTGCAAGAGTCTTCAAGATCCCTTTGTATACACATGAGGTTCAACATGG 544
 565 CTACGCAATTTGAAAGACTAGTGGAGACTCTGTCTCCCTCTCTCTCTCTCTCTCTCT 624
 545 CTACTACTAGAGAAACCTGTGGCCAGCTTCCATTCCTCTCTCTCTCTCTCTCTCTCT 604
 625 TATGTCAAGAACTAAGTACCAATGATCTTTCATAGAGAGGTTAAATAATATGATCT 684
 605 CATGAGTGAAGTGAACAAATATGACCTTTCACAGAGAGGTTGAAAAATATGATGTTTC 664
 685 GCTTATTTTGAATTTTGTCTCCAAATATTGACATGAAAGAGTGGATCACTTCTACAG 744
 665 ACTGTGTGTGAGTACTGTGCTCCAGCAATATGACTTTGCAATTCGGGATCAGTTTACAG 724
 745 TGAAGTTCTAGAGACCACTACAGTTATCTGAGCAATGAGCAAAAGTGCATATGAGCT 804
 725 TGAAGCTTGAAGAGCCCAAGAGTTTGTGAAGCTGTGGGAGAGTGCATTTGGCT 784
 805 TATTCGAAACTACTGGATTTTCAATTTCTACCCCACTTTCACCAATGTTGAGTTCT 864
 785 AATCCGAACATATTTGGAGTGTGAGTTTCTCTGTCATATTTACCAATTTTGGATTTGT 844
 865 TGGAGACTCCATGCAAACTGCGCAAACTCCCTTACCGAAGAAATGAAAGATTGTCGA 924
 845 GGGAGGACTGCACTGCAAACTGCGCAAACTCTTTCCTTAAAGAAATGAAAGATTGTTCA 904
 925 GAGCTTGGAGAAATGAGTGTGAGTTTCTCTGAGGCTGAGTGTGATACATACAGTCT 984
 905 GAGCTTGGAGAAATGAGTGTGAGTTTCTCTGAGGCTGAGTGTGATACATACAGTCT 964
 985 AGAAGAAAGGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1044
 965 AGAAGAAAGGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1024

QY 1045 GAGATTTGATGAGAAATTAACAGATACTTTAGAGCTCATACTGCTGGCTGACAGTGGAT 1104
 DB 1025 GAGATTTGATGAGAAATTAACAGATACTTTAGAGCTCATACTGCTGGCTGACAGTGGAT 1084
 QY 1105 ACCCGAATGATCTTCTTGTGATCAATCCCAAAACCAAGCTTTTATCATCATGCTGAGAT 1164
 DB 1085 TCCCGAATGATCTTCTTGTGATCAATCCCAAAACCAAGCTTTTATCATCATGCTGAGAT 1144
 QY 1165 GAATGGATCTATGAGATCTTATTAATCCATGGGCTCCCTATGTTGGAGTCCCATATTTGG 1224
 DB 1145 AAACGGATTTATGAAACCAATTAACATGGGCTCCCTATGTTGGAGTCCCATATTTGG 1204
 QY 1225 TGATGACTTGTATACATGATCTCAATGAAAGGCGCAAGAGAGCTGTAGAAATTAATCTT 1284
 DB 1205 GATCAGCTTCAACATGCTCTCATGAGGCGCAAGAGAGAGCTGTAGAAATTAATCTT 1264
 QY 1285 CAATATGATGACAGAGATTTACTGAGGCTTTGAGAAACAGTCAATACCAATTCCTC 1344
 DB 1265 CAGTACAAATGACAGAGATTTACTCAGTGTGTGAGGAGAGTCAATATGAGCTTC 1324
 QY 1345 TTATTAAGAAATGCTATGAGATTTATCAAGATTTACCATGATCACTGTAAAGCCCT 1404
 DB 1325 TTATTAAGAAATGCTATGAGATTTATCAAGATTTACCATGATCACTGTAAAGCCCT 1384
 QY 1405 AGATGAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1464
 DB 1385 GAGAGAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1444
 QY 1465 ATGAGCTGCGCAATGAGCTTCACTGATGATGATGATGATGATGATGATGATGATG 1524
 DB 1445 TGAGGAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1504
 QY 1525 GCTGACTGTGTGCAATGCTCTATATTTCTGTTCAAAAATGTTTATTTTCTGCTCA 1584
 DB 1505 ATGTTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1564
 QY 1585 AAAATTTAATTAAGTAAAGATGAGAAAGAGGAAATTA 1625
 DB 1565 AAAATTTAATTAAGTAAAGATGAGAAAGAGGAAATTA 1604

RESULT 8
 BC048920
 LOCUS 2575 bp mRNA linear HTC 27-MAR-2003
 DEFINITION Mus musculus, clone IMAGE:6311381, mRNA.
 ACCESSION BC048920
 VERSION BC048920.1 GI:29294723
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2575)
 Strausberg, R.
 Direct Submission
 Submitted (17-MAR-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-f@mail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Heltton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

Db	36	TTAAAGTACAGGAAGACCCACACACTGTGTGAGACTATGGGGAAGCTGAGATTGGACT	95
Qy	805	TATTCGAAACTACTGGGATTTTCAATTTCTCTACCCACTCTTACCAATGTTGAGTTGCT	864
Db	96	AATGCGAACCATTATGGGATTTTGAATTTCTCGTCTATATTACCGAATTTTGAAGTTGT	155
Qy	865	TGAGAGACTCCACTGCAAACTGGCCAAACCCCTACCGAAGGAATGGAAGTTGTGCCA	924
Db	156	GGGAGGACTGCACTGCAAACTGGCCAAACCCCTTACTAAGGAATGGAAGGTTGTGCCA	215
Qy	925	GAGCTCTGGAGAAAATGGTGTGTGTGTTTCTCTGGGGTCGATGCTCAACAACGTC	984
Db	216	GACCTCAGGGGAAACATGGTATTTGGTGTGTTTCTCTGGGGTCGATGCTCAACAACGTC	275
Qy	985	AGAGAAAAGGGCCCAATGTAATTTGATCAGCCCTTGCCAAAGTCCCAAAAAGTTCTGTG	1044
Db	276	AGATGAAAAGGCCAATCTCATTTGCTCAGCCCTTGCCCAAGTTCTCTCAAGAGTTTGTG	335
Qy	1045	GAGATTTGATGGGAAATAACAGATACTTTAGACTCAATACGCGGCTGTAACAAGTGAT	1104
Db	336	GCGATACAAAGGAAGATACAGACACACTTAGATCCCAATCAAGACTGTTGATTGGAT	395
Qy	1105	ACCCGAGAATATCTTCTTGTGTCATCCCAAAACCAAGCTTTTATCACTAGTGTGAAT	1164
Db	396	TCCCTCAGATATCTTCTTGGACATCCCAAAACAGAGCTTTTATCAACCAATGGTGAAAC	455
Qy	1165	GAATGGGATCTATGAAAGCTATTTTACATGGGGGTCCCTATGGTGGAGTTCCTATTTGG	1224
Db	456	AAATGGAATCTATGAGGCTATTTTACATGGGATCCCTATGGTGGAGTCCCATGTTGCG	515
Qy	1225	TGATCAGCTGANTAACATAGCTCACATGAAGGCCAAAGAGCAAGCTGTGAATAACTT	1284
Db	516	TGACCAGCCTACAAACATTTGCTCATATMAAGGCCAAAGAGCAGCAGTGGAGGTGAACAT	575
Qy	1285	CAAAACCTATGCAAGCGAAGATTTACTAGGGCTTTGGAACAGTCATTACCGATTCCTC	1344
Db	576	GAAACGATGACAAAGCTCAGATCTGCTCAATGCTCTGGAACAGTCATCAATGAGCCATC	635
Qy	1345	TTATTAAGAGATGCTATGAGTTATCAAGAATTCACCATGATCAACCTGTAAAGCCCT	1404
Db	636	TTATTAAGAGATGCTATGAGTTATCAAGAATTCACCATGATCAACCTGTAAAGCCCT	695
Qy	1405	AGATCGAGACGCTTCTGGAATCGAGTTGTCAATGGCCCAAAAGGAGCAAGCAAGCTGGC	1464
Db	696	GGACCGAACAATCTTCTGATTTGAATTTGTCAATGGTCAAAAGGAGCAAGCAAGCTGGC	755
Qy	1465	ATCAGCTCCCATGACCTCACCTGGTTCCAGACACTCTATAGATGTGATTTGGTTCT	1524
Db	756	TGTGCAAGCCCATGACCTCAAGTGTTCAGTACACACTCTCTGAGATGTAATTTGGTTCT	815
Qy	1525	GCTGACCTGTGTGCAACTGCTAATCTTGTTCACAAAATGTTTTTAATTT-TCTGTCT	1584
Db	816	GCTGGGCTGTGTGCGCATCTGCTATATTTATTTGGTTGCAAAAATGCTGTGTTTATTAATTC	875
Qy	1584	AAAAATTTAATTAACGTGAAGAAGATAGAAAAGAGGAAATAGAT-1626	
Db	876	AAAAAGTTGTAGACAGAGAAAGGAAAAAAGCTGACTAGTT-918	

RESULT	11
LOCUS	BQ942104
DEFINITION	BQ942104 1004 bp mRNA linear EST 21-AUG-2002 AGENCOURT_8763113 NIH_MGC_129 Mus musculus cDNA IMAGE:631334
ACCESSION	BQ942104
VERSION	BQ942104
KEYWORDS	5', mRNA sequence.
SOURCE	BQ942104.1 GI:22357582 EST.
ORGANISM	Mus musculus (house mouse)
	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1004) NIH-MGC http://mgc.nci.nih.gov/ .
REFERENCE	
AUTHORS	

```

TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: egadps-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: L1AM13740 row: f column: 05
High quality sequence stop: 656.
Location/Qualifiers
1. 1004
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6313348"
/lab host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 129"
/note="Organ: Olfactory epithelium; Vector:
pCMW-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unc directionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: This
is a NIH MGC library."
BASE COUNT
286 a 219 c 249 g 249 t 1 others
ORIGIN
Query Match 36.4%; Score 599.8; DB 13; Length 1004;
Best Local Similarity 80.0%; Pred. No. 6.2e-116;
Matches 718; Conservative 0; Mismatches 178; Indels 2; Gaps 1.

```

QY	721	GAAGAAGGGGAGTACAGTTCTACACGTGAAGTCTGTGAAGAACCACCTACAGTTATCTGAAC	780
Db	46	GAACACGTGGGATTCCTATTATATACATAAAGCTTTAGAAAGCCACACACACTGTGGAGAC	105
QY	781	AATGGCAAAAGCTGACATATGGCTTATTTCCAAACTACTGGGATTTTCAATTTCTCCACCC	840
Db	106	TATGGGGAAAGCTGAGATTTGGCTTAATGCCAACCTATTTGGGATTTTGAATTTCTCGTCC	165
QY	841	ACTCTTACCAATGTGAGTCTGTGAGGAGCTCCACTGCAACCTGCAAAACCTGCACACCCCTAAC	900
Db	166	ATATTTCAGGAATTTTGAGTTTGAGGGAGGAGCTGACTGCAAACTGTCCAAAGCCTTTTACC	225
QY	901	GAAGGAATATGAAAGTTTGTCCAGAGCTCTGAGAAATAGTGTTGTGTGTTTCTCT	960
Db	226	TAAAGAAATGAAAGATTTGTCTCAGAGCTCAGGGGGAACATGTATTTGTGGATTTTCTCT	285
QY	961	GGGGTCAGTGTCACTAACACGCTCAGAGAAAGAGGCCCAATGTATTTGCATCAGCCCTTGC	1022
Db	286	GGGGTCATGTGTCAAAAACCTGACAGATGAAGAAAGGCCAATCTCATTTGCTCAGCCCTTGC	345
QY	1021	CAAGATCCCAAAAAGTTCTGTGTGAATTTGATGGGAATTAACAGATCTTTAGSACT	1082
Db	346	CCAGATTCCTCAGAGAGTTTGTGGCCGATACAAAGAAAGATPACAGACACTTAGSACT	405
QY	1081	CAATACCTGGGCTGTACAAGTGSATACCCCAATATATCTTTGGTGTCAATCCCAAAACCA	1142
Db	406	CAATACCAAGACTGTTTGATGTGATTTCTCTCAGAAATATCTTTGTGACATCCCAAAACAG	465
QY	1141	AGCTTTTATCACTCATGTGTGAAATGAATGGGATCTATGAACTATTTTACCATGGGGTCCC	1202
Db	466	AGCTTTTATCACTCATGTGTGAAATGAATGGGATCTATGAACTATTTTACCATGGGGTCCC	525
QY	1201	TATGTGGGAGTCCCATATTTTGGTGTGATCAGCTTGATTAACATAGTCACTGAAAGGCCAA	1262
Db	526	TATGTGTGGAGATGCCCATGTTCGTGACAGAGCTGACAAACATGTCTCACTGAAAGGCCAA	585
QY	1261	AGGAGCACTGTAGAAATATAAATTCMAAACTATGACAGCGGAAGATTACTAGGGGCTTT	1322
Db	586	AGGAGCACTGTAGAGGTGAACATGAAACAGATGACAAAGCTTCAGATCTGTCTCAATGTCTT	645

QY	1321	GAGAAAGCTGCTTACCCGATTCCTCTTATTAAGAGATGCTTATGATATCAAGATTCA	1380
Db	646	GAGAAAGCTGCTTACCATGAGCCATCTTATTAAGAGATGCTTATGATATCAAGATTCA	705
QY	1381	CCATGATCAACCTGTAAAGCCCTTAGATCGAGCACTCTTCGTGATCGAGTTTGTTCATCG	1440
Db	706	CCATGATCAACCTGTAAAGCCCTTAGATCGAGCACTCTTCGTGATCGAGTTTGTTCATCG	765
QY	1441	CCACAAAGAGCCCAAGACCTCTGCGATCAGCTGCGCCATTAACCTACCTGCTTCCAGACTA	1500
Db	766	TCACAAAGAGCCCAAGACCTCTGCGATCAGCTGCGCCATTAACCTACCTGCTTCCAGACTA	825
QY	1501	CTCTATAGATGATGATTGGGTTCCCTGCTGACCCGTGTGGCAACGTGATATTCCT--GTTG	1558
Db	826	CTCTCTGAATGGAATTTGGGGTTCTCTGCTGCGCTGTGTGCAATCTGCTTAATTTATTTGGGTG	885
QY	1559	ACAAATGTTTTTTTATTTTCTGTCAAAATTTTAAATAACATAGACATGAAAAAGA	1616
Db	886	CAAAAGGTGGTGGTTTAAATTTTCAAAAAGTTGTTAGAACCGGAAAAAGAAAAAAA	943
RESULT 12			
BGS62901		595 bp	mRNA
LOCUS			
DEFINITION	602581752p1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4709436 5',		
ACCESSION	BGS62901		
VERSION	BGS62901.1	GI:13570553	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia, Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 595)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMD at: http://image.llnl.gov Plate: L10CM1546 row: h column: 13 High quality sequence stop: 595. Location/Qualifiers		
FEATURES			
source	1..595		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4709436"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/clone_lib="NIH MGC 76"		
	/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgccctcgccg); Site 2: SfiI (ggccatctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTAAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dT (30) BN-3' (where B = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."		
BASE COUNT	175 a 115 c 128 g 177 t		
ORIGIN			
Query Match	35.2%	Score 580.8; DB 10; Length 595;	
Best Local Similarity	99.7%	Prod. No. 6.1e-112;	
Matches 582; Conservative	0; Mismatches 2;	Indels 0; Gaps 0;	

OY	1	AGCAACTGGAAAA	CAGGATTTGCA	TTGGATGAGATGCTA	TGAAATGGA	CTTCAAGCT	60
Db	12	AGCAACTGGAAAA	CAGGATTTGCA	TTGGATGAGATGCTA	TGAAATGGA	CTTCAAGCT	71
OY	61	TTCTCCTGATAC	GAGCTGAGCTGT	TACTTACCTTGAG	CGAGTGTGGAAAGG	TGCTGCTG-120	
Db	72	TTCTCCTGATAC	GAGCTGAGCTGT	TACTTACCTTGAG	CGAGTGTGGAAAGG	TGCTGCTG-131	
OY	121	GCCCAAGAATTC	GACCACTGGA	TGAATATTAAGA	CAATCTTGATGA	ACTTGTCCAGAG	180
Db	132	GCCCAAGAATTC	GACCACTGGA	TGAATATTAAGA	CAATCTTGATGA	ACTTGTCCAGAG	191
OY	181	AGGTCATGAGGT	GACTGATTTGG	ATGCGATCTGCA	TTCTCTTGCGATCC	CAACAGGCC	240
Db	192	AGGTCATGAGGT	GACTGATTTGG	ATGCGATCTGCA	TTCTCTTGCGATCC	CAACAGGCC	251
OY	241	ATCTACTCTTAA	ATTGAAGTTTAT	TCCTGTATCTTTAA	CTAAAACTGAG	TGTGAGATAT	300
Db	252	ATCTACTCTTAA	ATTGAAGTTTAT	TCCTGTATCTTTAA	CTAAAACTGAG	TGTGAGATAT	311
OY	301	TATCAAGCAGCT	GGTGAAGATGG	GCGAAGCTTCC	AAAGACACATTTG	GTCTATATT	360
Db	312	TATCAAGCAGCT	GGTGAAGATGG	GCGAAGCTTCC	AAAGACACATTTG	GTCTATATT	371
OY	361	TTCAAGAATCA	AAATCATGTG	SACATTAATGACA	TACTTTGAAG	AGTCTGTAAAGA	420
Db	372	TTCAAGAATCA	AAATCATGTG	SACATTAATGACA	TACTTTGAAG	AGTCTGTAAAGA	431
OY	421	TATAGTTCAAA	TATAGAAACTT	ATGAAAGAACTA	CAGAGTCAAG	ATTGATGTTGCT	480
Db	432	TATAGTTCAAA	TATAGAAACTT	ATGAAAGAACTA	CAGAGTCAAG	ATTGATGTTGCT	491
OY	481	TGCAATGCTGTT	TTCCCTTTGGT	GAGCTGCTG	CGAGTTACTTAA	AAATACCTTTGT	540
Db	492	TGCAATGCTGTT	TTCCCTTTGGT	GAGCTGCTG	CGAGTTACTTAA	AAATACCTTTGT	551
OY	541	CTAAGGCTCGG	CTCTCTCGGCT	ACGCAATTTGA	AAAGATA		584
Db	552	CTAAGGCTCGG	CTCTCTCGGCT	ACGCAATTTGA	AAAGATA		595

```

RESULT 13
BG428781
LOCUS
DEFINITION
      BG428781 814 bp mRNA linear EST 14-MAR-2001
      60250086CP1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4614566 5',
      mRNA sequence.
ACCESSION
      BG428781
VERSION
      BG428781.1 GI:13335287
KEYWORDS
      EST.
SOURCE
      Homo sapiens (human)
ORGANISM
      Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 814)
      NIH-MGC http://mgc.nci.nih.gov/.
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished
      Contact: Robert Strausberg, Ph.D.
      Email: cgabbs@email.nih.gov
      Tissue Procurement: CLONTECH Laboratories, Inc.
      cDNA Library Preparation: CLONTECH Laboratories, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
      DNA Sequencing by: Incyte Genomics, Inc.
      Gene distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LNLN at:
      http://image.llnl.gov
      Plate: LNCM1366 row: 9 column: 15
      High quality sequence stop: 761.
      Location/Qualifiers
          1..814
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
FEATURES
      SOURCE

```

```

/clone="IMAGE:4614566"
/lab host="DH10B (T1 phage-resistant)"
/clone.lib="NIH_MGC_75"
/notes="Organ: kidney; Vector: pDNr-LIB (Clontech); Site_1:
5'fl (ggcgcgcgcgcgc); Site_2: 5'fl (ggcgcgcgcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGCGCATGCG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT      224 a      167 c      182 g      241 t
ORIGIN
Query Match      34.9%; Score 576.4; DB 10; Length 814;
Best Local Similarity 85.8%; Pred. No. 5.3e-111;
Matches 688; Conservative 0; Mismatches 106; Indels 8; Gaps 4;
QY 453 TACAGAGTCAAGATTGATGTTCTTTCAGAGATGCTTTCCCTTGGTGGAGCTGC 512
Db 1 TACAAGAGTCAAGATTGATGTTCTTTCAGAGATGCTTTTCCTGTAGTGGAGCTGC 60
QY 513 TGGCCGAGTACTTAAATACCTTTGTCTACAGCGCTTCCTCTCGGCTACGCA 572
Db 61 TGGCTGAGCTATTAACTACCTTTGTGTACAGTCTCAGCTTCTCTGCTACATT 120
QY 573 TTGAAGACATAGTGAAGACTTCTGTTCCCTCTCTTCTTCTTCTTCTTCTTCT 632
Db 121 TTGAAGACATAGTGAAGACTTATTTTCCCTCTCTCTCTCTCTCTCTCTCTCT 180
QY 633 AACTAGAGCAACCAATGACTTCTTATGAGAGGCTTAAATATGATGATGCTTANT 692
Db 181 AATTAAGTATCAATGACTTCTTATGAGAGGCTTAAATATGATGATGCTTANT 240
QY 693 TTGAATTTTGGTCCAAATATTTGACATGAGAGAGTGGATCAGTTCTTCAAGTAA 752
Db 241 TTGACTTTTGGTCCAAATATTTGACATGAGAGAGTGGATCAGTTCTTCAAGTAA 300
QY 753 TAGAAGAGCCACTACGTTATCTGAGACATGAGCAAAAGCTGACATATGCTTAT 812
Db 301 TAGAAGAGCCACTACATATCTGAGACATGAGCAAAAGCTGACATATGCTTAT 360
QY 813 ACTATCTGGATTTTCAATTTCTCCACCTCTTACCAATGTTAGTTCTGTTAGAG 872
Db 361 ACTCTGGAATTTTCAATTTCTCCACCTCTTACCAATGTTAGTTCTGTTAGAG 419
QY 873 TCCACTGCAAACTGCAAAACCTTACCGAAGAAATGGAAGTTTGTCCAGAGCTGT 932
Db 420 TCCACTGCAAACTGCAAAACCTTGTCCGAGAAATGGAAGTTTGTGACAGAGCTGT 479
QY 933 GAG-AAAATGTTGTTGTTGTTTCTCTGAGGTCGATGTCAGTAACTGCAAGAA 991
Db 480 GAGCAAAAATGTTGTTGTTGTTTCTCTGAGGTCGATGTCAGTAACTGCAAGAA 539
QY 992 AGGCGCAATGTAATGTCATGACGCTTGGCAAGATCCCAAAAGTTCTGAGATTT 1051
Db 540 AGGCGCAATGTAATGTCATGACGCTTGGCAAGATCCCAAAAGTTCTGAGATTT 598
QY 1052 GATGGGAATTAACAGATCTTATAGACTCAATCTGCGCTGTAACAAGTATACCC 1111
Db 599 GATGGGAATTAACAGATCTTATAGACTCAATCTGCGCTGTAACAAGTATACCC 658
QY 1112 AATGATCTTCTTGTGTCATCCCAAAACCAAGCTTTTATCACTCATGTTGAATG 1166
Db 659 AATGATCTTCTTGTGTCATCCCAAAACCAAGCTTTTATCACTCATGTTGAATG 718
QY 1167 ATGGATCTATGAGATTTTACCATGGGCTCCATGTTGGAGAGTTCCCATATTTGG 1226
Db 719 TGGGATCTATGAGAGATTTTACCATGGGCTCCATGTTGGAGAGTTCCCATATTT 778
QY 1227 ATCAGCTGATTAACATAGCTCA 1248

```

```

Db 779 ATCAACCTGTATAATGCTCA 800
|||||
RESULT 14
B0925596 927 bp mRNA linear EST 20-AUG-2002
LOCUS
DEFINITION
AGENCOURT 8688479 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311424
5', mRNA sequence.
B0925596
ACCESSION
B0925596.1 GI:22340627
VERSION
B0925596.1
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 927)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: rgabs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM13735 row: E column: 01
High quality sequence stop: 635.
Location/Qualifiers
1..927
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6311424"
/lab host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_129"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH MGC Library."
BASE COUNT      262 a      201 c      230 g      231 t      3 others
ORIGIN
Query Match      34.9%; Score 575.4; DB 13; Length 927;
Best Local Similarity 78.3%; Pred. No. 8.6e-111;
Matches 702; Conservative 0; Mismatches 193; Indels 2; Gaps 1;
QY 745 TGAAGTCTAGAGAGCCCACTAGCTATCTGAGACAAATGGAAGCTGACATATG 804
Db 28 TAAAGTACAGAGAGCCCACTAGCTATCTGAGACAAATGGAAGCTGACATATG 87
QY 805 TATTCGAAATCTAGGATTTTCAATTTCTCACCCTCTTACCAATGTTGATTCG 864
Db 88 AATGCAATCTATGGAATTTTGAATTTCTCCTGTCATATTTACGAATTTGATTCG 147
QY 865 TGAAGACTCAGTCAAACTGCAAAACCTTACCGAAGAAATGGAAGTTTGTTC 924
Db 148 GGAAGACTCAGTCAAACTGCAAAACCTTACCGAAGAAATGGAAGTTTGTTC 207
QY 925 GAGCTCTGAGAAATATGTTGTTGTTTCTCTGAGGTCGATGTCATACAGTCT 984
Db 208 GACCTCAGGGAATATGTTGTTGTTTCTCTGAGGTCGATGTCATACAGTCT 267
QY 985 AGAAGAAAGGCAATGTAATTTGATCAGCCCTTGCAAGATCCCAAAAGTTCTGTG 1044
Db 268 AGATGAAGAGCAATCTCATTTGCTCAGCCCTTGCAAGATTTCTCAAGAGTTTGTG 327
QY 1045 GAGATTTGATGGAATTAACAGATCTTATAGACTCAATCTGCGCTGTAACAAGT 1104

```

```

Db      328 GCGATCAAGAGAAAGATACAGACATTAAGATCCAAATCAAGACTGTTTGAATGAT 387
Qy      1105 ACCCCAGATGATCTTCTTGGTCAATCCCAAAACCAAGCTTTTATCACTCATGATGAGAT 1164
Db      388 TCTTCAGATATATCTTCTTGGACATCCAAAACCAAGCTTTTATCACTCATGATGAGAT 447
Qy      1165 GAATGGAGTCTATGAAGCTATTTACCATGGGGTCCCTATGATGGAGAGTTCCCATATTTGG 1224
Db      448 AATATGATCTATGAGGCTATTTACCATGGAGTCCCTATGATGGAGAGTCCCATGTTGCG 507
Qy      1225 TGATCAGCTTATTAACATAGCTACATGAGGCCCAAGAGACAGCTGTAGAAATTAACCTT 1284
Db      508 TGACAGGCTTACCAACATTTGCTACATGAGGCCCAAGAGACAGAGAGAGTGAAGTGAACAT 567
Qy      1285 CAAAACTATGACCAAGCAAGATTTACTGAGGGCTTTGAGAACATGATTCAGATTCCTC 1344
Db      568 GAACACGATGACCAAGCTCAGATCTGCTCAATGCTTGAAGACATGATCAATGAGCCATC 627
Qy      1345 TTATTAAGAGATGCTATGAGATTTACCAAGATTCACCATGATCAACCTGTAAAGCCCT 1404
Db      628 TTATTAAGAGATGCTATGAGATTTACCAAGATTCACCATGATCAACCTGTAAAGCCCT 687
Qy      1405 AGATCGACAGCTCTTCTGATGAGTTGTGATGCGCCCAAGAGACCAAGCACTGCG 1464
Db      688 GGACCGAGCAGCTCTTCTGATGAGTTGTGATGCGCCCAAGAGACCAAGCACTGCG 747
Qy      1465 ATCAGCTGCGCAGCTCAGCTGCTGACAGCACTCTATGATGATGATGGTCT 1524
Db      748 TGTGCGACCGCAGCTCAGCTGCTGACAGCACTCTGATGATGATGATGGTCT 807
Qy      1525 GCTGACCTGCTGCGCACTGCTATATTTCTT--GTTCACAAATGTTTTTATTTTCTGT 1582
Db      808 GCTGCGCTGCTGCGCACTGCTATATTTCTTGGTTGCCAAAGGGGTTTCCGTTAAAT 867
Qy      1583 CAAAATTTAATAAATACTAGAAAGATAGAAAAGAGGATGATCTTCCAAATTC 1639
Db      868 TTCAAAAGATTTGGTTAGAACCGGGGAGAAAAGAAAACCGTGGACCTTAA 924

RESULT 15
HSM087819 standard; RNA; EST; 588 BP.
XX      AC      BX494989;
XX      SV      BX494989.1
XX      DT      09-MAY-2003 (Rel. 75, Created)
XX      DT      09-MAY-2003 (Rel. 75, Last updated, Version 1)
DE      Homo sapiens mRNA; EST DKFZp779L0816_r1 (from clone DKFZp779L0816)
XX      EST; expressed sequence tag.
XX      Homo sapiens (human)
XX      OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX      Eutheria; Primates; Catarrhini; Homiidae; Homo.
XX      RN      [1]
XX      RP      1-588
RA      Mambuti R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA      Han M., Wiemann S.;
RL      Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL      MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX      This is the 5' sequence of the clone insert
CC      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC      Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC      Sequenced by AGOHA (Berlin/Germany) within the cDNA sequencing
CC      Consortium of the German Genome Project.
CC      No 5' sequence available.
CC      This clone (DKFZp779L0816) is available at the RZPD in Berlin.

```

```

CC      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC      14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source
FT      1. 588
FT      /db_xref="taxon:9606"
FT      /mol_type="mRNA"
FT      /organism="Homo sapiens"
FT      /clone="DKFZp779L0816"
FT      /clone_1ib="779 (synonym: hnccl). Vector pSport1_Sfi; host
FT      DH10B; sites SfiIA + SfiIB"
FT      /dev_stage="fetal"
FT      /tissue_type="liver"
XX      SQ      Sequence 588 BP; 170 A; 114 C; 127 G; 177 T; 0 other:
XX      Query Match      34.8%; Score 573.4; DB 2; Length 588;
XX      Best Local Similarity 99.8%; Pred. No. 2.2e-110;
XX      Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AGCACTGGAAGAAACAGCATTTGATTCATGAGATGCTATGAAATGAGACTTCACTCT 60
Db      14 AGCAACTGGAAGAAACAGCATTTGATTCATGAGATGCTATGAAATGAGACTTCACTCT 73
Qy      61 TCTCTGATACAGCTGAGCTGTTACTTACTTACCTGCGGAGTTGTGAAGGCTGTGTG 120
Db      74 TCTGCTATACAGCTGAGCTGTTACTTACTTACCTGCGGAGTTGTGAAGGCTGTGTG 133
Qy      121 GCCCAGAAATTCAGCCACTGATGATTAAGAAACAAATCCTGATGAACCTTGTCAGAG 180
Db      134 GCCCAGAAATTCAGCCACTGATGATTAAGAAACAAATCCTGATGAACCTTGTCAGAG 193
Qy      181 AGGTATGAGAGTGAATGATTTGGCATCTTCACTTCTTCTTCCATCCCAACAGCCC 240
Db      194 AGGTATGAGAGTGAATGATTTGGCATCTTCACTTCTTCTTCCATCCCAACAGCCC 253
Qy      241 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTTAACTGAAGTTGAGATAT 300
Db      254 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTTAACTGAAGTTGAGATAT 313
Qy      301 TATCAGCAGCTGTTAAGATGAGAGAGCAACTTCCAAAAGACACATTTGGTCATATTT 360
Db      314 TATCAGCAGCTGTTAAGATGAGAGAGCAACTTCCAAAAGACACATTTGGTCATATTT 373
Qy      361 TTCAAGTACAGAAATCATGAGATTTAATGATTAATGATTAATGATTAATGATTAATGATTA 420
Db      374 TTCAAGTACAGAAATCATGAGATTTAATGATTAATGATTAATGATTAATGATTAATGATTA 433
Qy      421 TATAGTTTCAAAATGAGAACTTATGAGAAACTACAGAGTCAAGATTTGATGTTGTTCT 480
Db      434 TATAGTTTCAAAATGAGAACTTATGAGAAACTACAGAGTCAAGATTTGATGTTGTTCT 493
Qy      481 TGCAATGCTGTTTCCCTTTGGTGAAGTGTGCGGAGTTACTTAAATACCTTTGT 540
Db      494 TGCAATGCTGTTTCCCTTTGGTGAAGTGTGCGGAGTTACTTAAATACCTTTGT 553
Qy      541 CTACAGCTTCGCTTCTTCTGCTGCTAGCAATTG 575
Db      554 CTACAGCTTCGCTTCTTCTGCTAGCAATTG 588

```

Search completed: December 7, 2003, 07:38:01
 Job time : 2473 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)